apertura Volumen 18, número 2, pp. 58-73 | Universidad de Guadalajara
http://doi.org/10.32870/Ap.v18n2.2858

Evaluación del pensamiento algorítmico: análisis psicométrico de diferentes formas de evaluación en un estudio piloto

Assessing Algorithmic Thinking: A Psychometric Analysis of Different Assessment Approaches in a Pilot Study

Ismael Esquivel-Gámez*
Universidad Veracruzana, México
https://orcid.org/0000-0001-7914-5170

Martín Guerrero Posadas**
TecNM Campus San Luis Potosí, México
https://orcid.org/0000-0001-8635-3074

Nancy Araceli Olivares Ruíz***
Universidad Veracruzana, México
http://orcid.org/0009-0005-6005-9332

Recepción del artículo: 27/03/2026 | Aceptación para publicación: 17/09/2026 | Publicación: 30/09/2026

RESUMEN

El pensamiento algorítmico constituye una habilidad clave en la educación superior; sin embargo, su evaluación ha sido limitada al centrarse en resultados finales más que en procesos cognitivos. Este estudio tuvo como objetivo analizar el comportamiento psicométrico de un instrumento para evaluar el pensamiento algorítmico, mediante diferentes formas de evaluación del desempeño en estudiantes universitarios mexicanos, con diferentes niveles de habilidad. Se diseñó y aplicó una prueba de rendimiento máximo compuesta por cuatro formas de evaluación a 141 estudiantes universitarios de dos instituciones públicas, utilizando indicadores de la teoría clásica de los tests y análisis cualitativo de percepciones. Los resultados mostraron que el funcionamiento de las diversas formas dependió del nivel de habilidad y evidenció discrepancias entre la dificultad percibida y el desempeño observado. Como limitación, se reconoce el carácter de pilotaje y el tamaño reducido de los grupos. El estudio aporta evidencia sobre la necesidad de enfoques diferenciados de evaluación. Se concluye que la evaluación del pensamiento algorítmico debe ajustarse al nivel formativo.

ABSTRACT

Algorithmic thinking is a key skill in higher education; however, its assessment has been limited by a predominant focus on final outcomes rather than underlying cognitive processes. This study aimed to analyze the psychometric performance of an instrument designed to assess algorithmic thinking through different performance assessment approaches in Mexican university students with varying levels of ability. A maximum-performance test comprising four assessment approaches was developed and administered to 141 students from two public institutions, using indicators from classical test theory and a qualitative analysis of participants' perceptions. The results showed that the functioning of the assessment approach depended on students' ability levels and revealed discrepancies between perceived difficulty and observed performance. As a limitation, the pilot nature of the study and the relatively small group sizes are acknowledged. The study provides evidence supporting the need for differentiated assessment approaches. It is concluded that the assessment of algorithmic thinking should be aligned with students' level of development.

Palabras clave
Pensamiento algorítmico; evaluación educativa; análisis psicométrico; formas de evaluación

Keywords
Algorithmic thinking; Educational assessment; Psychometric analysis; Assessment approaches

SOBRE LOS AUTORES

* Doctor en Tecnología de Información y Análisis de Decisiones. Profesor de tiempo completo de la Universidad Veracruzana, México. ORCID: https://orcid.org/0000-0001-7914-5170, correo electrónico: iesquivel@uv.mx

** Doctor en Educación por la Nova Southeastern University. Profesor investigador del TecNM Campus San Luis Potosí, México. ORCID: https://orcid.org/0000-0001-8635-3074, correo electrónico: martin.gp@slp.tecnm.mx

*** Maestra en Sistemas de Información por la Fundación Arturo Rosenblueth de la Ciudad de México. Profesora de tiempo completo de la Universidad Veracruzana, México. ORCID: http://orcid.org/0009-0005-6005-9332, correo electrónico: nolivares@uv.mx

INTRODUCCIÓN

El pensamiento algorítmico (PA) se ha consolidado como una habilidad clave para la resolución estructurada de problemas en la educación superior, particularmente en áreas STEM, al implicar la formulación de soluciones mediante procedimientos secuenciales y sistemáticos (Wing, 2006). No obstante, su desarrollo representa un desafío para muchos estudiantes debido a las demandas cognitivas asociadas a la abstracción, la secuenciación y la comprensión de estructuras algorítmicas (Lister *et al*., 2004).

A pesar de su relevancia, la evaluación del PA aún es limitada, ya que las aproximaciones tradicionales se centran en los resultados y no en los procesos cognitivos implicados en la construcción de soluciones (Brennan & Resnick, 2012). En este contexto, el pensamiento algorítmico puede valorarse mediante diferentes formas de evaluación, entendidas en este estudio como maneras específicas de solicitar al estudiante que evidencie su desempeño algorítmico a través de un determinado tipo de tarea. Entre los antecedentes documentados en educación en programación se encuentran la lectura de código y la predicción de resultados –vinculadas con tareas de trazado de código (*code tracing*)–, así como el completado de algoritmos y el ordenamiento de pasos, este último formalizado en los denominados Parsons problems (Lister *et al*., 2004; Parsons & Haden, 2006; Denny *et al*., 2008).

La incorporación de diferentes formas de evaluación permite examinar el desempeño algorítmico ante tareas distintas, evitando circunscribir su evaluación a un único tipo de actividad. En este sentido, persisten vacíos en la validación de instrumentos que integren diversas formas de evaluación del pensamiento algorítmico, especialmente en su comportamiento en estudiantes con distintos niveles de experiencia. Debido a que el PA no se desarrolla de manera homogénea y su adquisición puede resultar compleja en etapas iniciales, resulta pertinente analizar su desempeño en función del nivel de exposición a contenidos algorítmicos en contextos educativos mexicanos, caracterizados por grupos reducidos y heterogéneos. En este marco, el presente estudio tiene como objetivo analizar el comportamiento psicométrico de un instrumento para evaluar el pensamiento algorítmico en estudiantes universitarios, considerando distintas formas de realizarlo

El fortalecimiento del pensamiento algorítmico contribuye al desarrollo del pensamiento computacional, al proporcionar las bases procedimentales necesarias para la construcción de soluciones en distintos niveles de complejidad

y niveles de experiencia, mediante indicadores derivados de la teoría clásica de los *tests*.

Pensamiento computacional y pensamiento algorítmico

El pensamiento computacional (PC) se ha consolidado como un constructo clave para la resolución estructurada de problemas, al implicar procesos orientados a diseñar soluciones sistemáticas basadas en principios de la informática, sin requerir necesariamente el uso de tecnologías digitales (Wing, 2006, 2008). En el ámbito educativo, se reconoce como una competencia relevante por su relación con el razonamiento lógico, la capacidad analítica y la resolución de problemas (Guitert y Romeu, 2022; Méndez y Bermúdez, 2023). Sus antecedentes se remontan al pensamiento procedimental propuesto por Papert (1980) y su desarrollo conceptual ha sido impulsado por iniciativas orientadas a su integración en contextos educativos, así como por marcos para su evaluación (Barr & Stephenson, 2011; Brennan & Resnick, 2012).

Dentro de este marco, el PA constituye un componente fundamental del PC, centrado en la formulación de procedimientos estructurados para la resolución de problemas. Se define como la capacidad de descomponer un objetivo en una secuencia ordenada de pasos ejecutables de manera sistemática, implicando habilidades como la abstracción, la planificación y la automatización (Pinzón y González, 2022). A diferencia del PC, que engloba un conjunto amplio de procesos cognitivos, el PA se enfoca específicamente en la construcción de soluciones paso a paso, lo que lo posiciona como un elemento clave para el análisis y diseño de algoritmos (Navas, 2024). En contextos educativos, el desarrollo del PA resulta especialmente relevante en la educación superior, donde se requiere que los estudiantes analicen problemas complejos y generen soluciones reproducibles. Su importancia radica en que favorece el razonamiento lógico y la resolución estructurada de problemas, tanto en disciplinas como la programación y las matemáticas, al igual que en otros contextos donde se demandan procesos sistemáticos de solución (Montero, 2024; Adorni y Piatti, 2025).

En este sentido, el fortalecimiento del PA contribuye al desarrollo del pensamiento computacional, al proporcionar las bases procedimentales necesarias para la construcción de soluciones en distintos niveles de complejidad.

Evaluación del pensamiento algorítmico

El creciente interés por el desarrollo del PA en la educación ha generado la necesidad de diseñar instrumentos que permitan evaluar estas habilidades de manera sistemática. Entre estos instrumentos se encuentra el *Web-Based Surveying Instrument to Identify Problem-Solving Abilities*, cuya estructura incluye ítems de opción múltiple, de completar y de resolución práctica, vinculados a trazado de código y ordenamiento de pasos (Vasconcelos-Santillan, 2008). Asimismo, el Test de Pensamiento Algorítmico en Adultos (ATTA) se compone de 20 ítems organizados en torno a algoritmos y representaciones, incorporando tareas de trazado de código y ordenamiento

de pasos (Lafuente *et al*., 2022). Por su parte, el Test de pensamiento algorítmico integra actividades provenientes de plataformas como Blockly Games y Lightbot, combinadas con tareas de trazado de código y ordenamiento de pasos (Posso y Murcia, 2022). Finalmente, el cuestionario para medir el pensamiento algorítmico se centra en la construcción y ejecución de algoritmos en contextos matemáticos, privilegiando tareas orientadas al trazado de código (Navas-López, 2022).

Estudios en México

En el contexto mexicano, la evaluación del PA se ha abordado principalmente como parte del pensamiento computacional, destacando su relación con habilidades como la resolución de problemas, el razonamiento lógico, la abstracción, la descomposición y la algoritmización, así como con otras áreas disciplinares (Gómez *et al*., 2025; Navas, 2024; Pinzón *et al*., 2023; Rodríguez-Abitia *et al*., 2021). Predomina un enfoque aplicado mediante intervenciones didácticas en educación básica y superior –incluyendo actividades lúdicas, programación, robótica y aprendizaje adaptativo– con resultados generalmente positivos (Rojas-López y García-Peñalvo, 2020). Entre los estudios que midieron el PA se encuentran:

Velasco-Ramírez y Otero-Escobar (2023) evaluaron el pensamiento algorítmico en estudiantes de segundo semestre de LSCA (n = 48) mediante el Test de pensamiento algorítmico (Posso y Murcia, 2022), en el contexto de una intervención tecnopedagógica. Los resultados mostraron mejoras en las habilidades de los estudiantes y evidenciaron que la colaboración entre pares favorece a quienes presentan menor nivel inicial. No obstante, el estudio se centró en los efectos de la intervención, ofreciendo información limitada sobre la medición específica de las dimensiones evaluadas.

González (2014) diseñó un instrumento basado en ejercicios de libros para evaluar el pensamiento algorítmico en estudiantes de primer semestre de Tecnologías de la Información (n = 22), en el marco de una intervención tecnopedagógica dirigida a alumnos interesados en videojuegos. El instrumento incluyó problemas de secuenciación, descomposición, ciclos y condicionales, evaluados mediante rúbrica. Sin embargo, los resultados se enfocan en ajustes del instrumento y efectos de la intervención, con información limitada sobre el comportamiento de las dimensiones evaluadas.

Velasco y Sánchez (2021) adaptaron el instrumento de Vasconcelos-Santillan (2008) para evaluar habilidades algorítmicas en estudiantes de LSCA de segundo semestre (n = 122), validándolo mediante juicio de expertos y coeficiente Kappa de Fleiss, y estimando su consistencia interna con KR-20 tras reducirlo a once ítems. Los resultados evidenciaron bajo nivel de competencia, especialmente en comprensión de instrucciones, manipulación algebraica y planificación algorítmica; no obstante, el instrumento careció de ítems para evaluar el completado de algoritmos.

MÉTODOS Y MATERIALES

El estudio corresponde a la fase de pilotaje de un instrumento para evaluar el pensamiento algorítmico en universitarios, cuyo objetivo fue analizar

En el contexto mexicano, la evaluación del PA se ha abordado principalmente como parte del pensamiento computacional, destacando su relación con habilidades como la resolución de problemas

el comportamiento psicométrico de los reactivos y el desempeño de distintas formas de evaluación, en grupos con diferentes niveles de experiencia. Se utilizó un diseño no experimental y transversal, con recolección de datos en un solo momento. Este enfoque permitió examinar el funcionamiento de los reactivos e integrar indicadores cuantitativos y valoraciones cualitativas obtenidas mediante cuatro preguntas abiertas.

Participantes

La muestra estuvo conformada por 141 estudiantes universitarios de dos IES públicas (IES = 1, sureste de México; IES = 2, centro del país). Se utilizó muestreo por cuotas con grupos intactos de primero, tercero, quinto y sexto semestre, con el fin de identificar el período evaluativo más adecuado. La edad promedio fue de 20.1 años y 42% correspondió a mujeres. Aunque los grupos presentaron tamaños desiguales, esto no representa una amenaza para la validez en estudios no experimentales centrados en análisis de ítems y en indicadores poco sensibles al tamaño muestral, como el índice de facilidad (Allen & Yen, 2002; Muñiz, 2010). Los participantes pertenecían a programas y semestres con distintos niveles de exposición a contenidos de algorítmica y fueron clasificados en cuatro niveles de experiencia algorítmica: (1) sin experiencia previa, (2) nivel inicial, (3) nivel intermedio y (4) nivel intermedio-avanzado, entendidos como una progresión funcional más que cronológica. La distribución completa se presenta en la tabla 1.

Instrumento

Se desarrolló el instrumento Carritos y Paquetes, una prueba de rendimiento máximo para evaluar el pensamiento algorítmico, integrada por cuatro formas de evaluación (Lectura de códigos, Resultados esperados, Código faltante y Ordenamiento de pasos), cada una con cinco reactivos de opción múltiple. Los ítems incluyeron mapas con casillas, un carrito que figuraba como el agente de ejecución y paquetes identificados mediante letras, junto con símbolos que indicaban posiciones específicas. Antes de la aplicación, se presentó un ejemplo con explicación, junto con instrucciones impresas y un demo en línea para facilitar la comprensión. El instrumento se imprimió con elementos ampliados para mejorar la legibilidad. Al finalizar, los estudiantes respondieron: (1) ¿cómo te sentiste?, (2) ¿cuál fue la más fácil?, (3) ¿cuál fue la más difícil? y (4) ¿qué mejoras propones para estos tipos de prueba?

Descripción de las formas de evaluación

El instrumento comprende tareas de lectura de código y resultados esperados, asociadas con actividades de trazado de código (Lister *et al.*, 2004; Denny *et al.*, 2008). Asimismo, se incluyeron ejercicios de código faltante, relacionados con tareas de producción parcial de código, y actividades de ordenamiento de pasos (Parsons & Haden, 2006) y posteriormente analizados por Denny *et al.* (2008) como una alternativa eficaz para evaluar la comprensión estructural de los algoritmos.

Tabla 1. Características de los grupos participantes según nivel de experiencia algorítmica

IES	Grupo	Nivel	Programa	Semestre	n	Edad media	Mujeres (%)
1	LSCA-3	3	LSCA	3º	47	20.7	44.7
1	LSCA-5	4	LSCA	5º	36	21.3	11.1
2	IND-6	1	II	6º	27	20.8	34.6
2	ISC-1	2	ISC	1º	31	18.4	25.8

Nota: II = Ingeniería Industrial; ISC = Ingeniería en Sistemas Computacionales; LSCA = Licenciatura en Sistemas Computacionales Administrativos.
Fuente: elaboración propia.

En términos operativos, en Lectura de códigos se presentaron cuatro grupos de instrucciones entre los cuales el estudiante debía seleccionar aquel que permitiera cumplir con el objetivo planteado; en Resultados esperados, se mostró un código completo y se solicitó anticipar el resultado final de su ejecución; en Código faltante, se requirió identificar el fragmento de código necesario para lograr un objetivo específico; y en Ordenamiento de pasos, el estudiante debió organizar correctamente una serie de acciones para alcanzar el resultado propuesto.

Se diseñaron cinco ítems por forma de evaluación con el fin de controlar la carga cognitiva y mantener la calidad de respuesta. En pruebas de rendimiento máximo, tiempos prolongados pueden generar fatiga y afectar el desempeño (Ackerman & Kanfer, 2009). Desde la Teoría de la Carga Cognitiva, la capacidad limitada de la memoria de trabajo exige regular la extensión de las tareas (Sweller, 2011). Además, se ha documentado que sesiones largas pueden propiciar desenganche del esfuerzo, afectando la validez de las respuestas (Wise, 2017).

Validez de contenido

Esta validez se estableció mediante juicio de expertos, entendida como el respaldo teórico y empírico de la interpretación de las puntuaciones (Almanasreh *et al*., 2019). Participaron trece profesores del área de sistemas computacionales, con edad media de 47.4 años y 19.8 años de experiencia docente; siete con doctorado y seis con maestría, con una distribución casi equitativa entre ambas instituciones. Evaluaron los ítems en claridad, relevancia, coherencia y suficiencia mediante una escala Likert de cinco puntos (1 = nada adecuado, 5 = muy adecuado). El acuerdo se estimó con el V de Aiken y sus intervalos de confianza al 95 %, considerando aceptables valores ≥ .70 (Merino, 2009).

Estudio piloto y aplicación

El instrumento se aplicó en un estudio piloto de seis semanas bajo un protocolo estandarizado. La prueba se realizó en formato físico y en dos sesiones para reducir la fatiga: la primera incluyó Lectura de códigos y Resultados esperados (30-40 minutos), y la segunda Código faltante y Ordenamiento de pasos (20-30 minutos). Aunque no hubo límite estricto de tiempo, los estudiantes debieron concluir dentro de una sesión de clase (dos horas). Previamente, se presentó un demo en línea para reforzar las instrucciones y el funcionamiento del instrumento.

En la validez de contenido participaron trece profesores del área de sistemas computacionales, con edad media de 47.4 años y 19.8 años de experiencia docente; siete con doctorado y seis con maestría

Procesamiento y análisis de datos

Debido a que el instrumento es una prueba de rendimiento máximo y no una escala tipo Likert, no se estimó el alfa de Cronbach. Además, al integrar cuatro formas de evaluación con demandas cognitivas diferenciadas y contar con pocos reactivos por cada una, se priorizó el análisis de ítems como estrategia principal (Crocker & Algina, 2008; Ebel & Frisbie, 1991). El análisis cuantitativo se centró en el índice de facilidad (IF), el índice de discriminación (D) y el coeficiente punto biserial (r_pb), lo que permitió identificar ítems adecuados y aquellos que requerían ajuste (Crocker & Algina, 2008; Haladyna & Downing, 1989).

Para las preguntas abiertas, se utilizó análisis de contenido inductivo, mediante codificación de unidades de significado y su agrupación en categorías interpretativas, incorporando procesos de contextualización y descontextualización para refinar los resultados (Elo & Kyngäs, 2008; Erlingsson & Brysiewicz, 2017; Bengtsson, 2016). El proceso fue revisado de forma iterativa para asegurar la consistencia del análisis.

Consideraciones éticas y consentimiento informado

La participación fue voluntaria y se obtuvo consentimiento informado por escrito. Se explicó el propósito académico del estudio y el uso exclusivo de los datos con fines de investigación. Para garantizar la confidencialidad, se asignaron claves numéricas en lugar de datos identificables. Asimismo, se aseguró que la participación no implicaba riesgos y que los resultados no tendrían efectos en la evaluación académica.

RESULTADOS

Validez de contenido

En términos generales, los reactivos presentaron valores adecuados del coeficiente V de Aiken en los cuatro criterios evaluados: claridad (.90), relevancia (.94), coherencia (.89) y suficiencia (.90). Únicamente un ítem correspondiente a Ordenamiento de pasos obtuvo un valor inferior al criterio establecido para claridad (V = .65), lo que motivó ajustes en su redacción antes de la aplicación del instrumento. Los resultados detallados por forma e ítem, así como los intervalos de confianza a 95%, se presentan en la tabla 2.

De manera complementaria, se solicitó a los expertos que ordenaran las formas de evaluación de acuerdo con la dificultad percibida. A partir de este juicio, se obtuvo la siguiente secuencia de menor a mayor complejidad: Lectura de códigos, Resultados esperados, Código faltante y Ordenamiento de pasos. Debido a que el instrumento se encuentra en fase de pilotaje, los análisis realizados se orientaron principalmente al comportamiento de los ítems y a la identificación de patrones de desempeño diferenciado.

Análisis psicométrico en fase de pilotaje

El cálculo de los índices psicométricos se realizó por grupos y por forma, para así analizar el comportamiento de los ítems según el nivel de experiencia algorítmica. Se estimaron el índice de facilidad (IF), el índice de discriminación (D) y el coeficiente punto biserial (r_pb), priorizando D como indicador principal por su mayor robustez, especialmente en muestras pequeñas. La r_pb se interpretó con cautela y se eliminaron ítems con correlaciones negativas, por su incongruencia con el constructo evaluado (Crocker y Algina, 1986; Ebel y Frisbie, 1991; Haladyna, 2016). El IF se utilizó como referencia complementaria para identificar valores extremos.

La selección de ítems se basó en criterios de la teoría clásica: se conservaron aquellos con D ≥ .30, y también se consideraron aceptables los ítems con D ≥ .20 y r_pb ≥ .20 en rangos intermedios de IF, particularmente en estudios exploratorios (Haladyna, 2016). Como se observa en la tabla 3, la retención de ítems no fue homogénea entre las formas ni entre los niveles de experiencia algorítmica, evidenciando una adecuación diferencial en su funcionamiento. Los casos con valores limítrofes o extremos fueron examinados de manera complementaria, sin constituir una categoría diferenciada en la tabla 3.

La forma de evaluación correspondiente a Lectura de códigos mostró un comportamiento relativamente estable, con una retención consistente de dos a tres ítems a lo largo de los niveles, mientras que Resultados esperados presentó el desempeño más inestable, con retención mínima o nula en varios niveles. Por su parte, Código

Tabla 2. Coeficiente V de Aiken por ítem y criterio (IC 95%)

Forma de evaluación	Ítem	Claridad (V)	Relevancia (V)	Coherencia (V)	Suficiencia (V)
Lectura de códigos	Q1	.92	.98	.94	.96
	Q2	.98	1.00	.96	.98
	Q3	.90	.92	.90	.90
	Q4	.94	.96	.92	.94
	Q5	.92	.98	.92	.96
Resultados esperados	Q1	.77	.96	.90	.90
	Q2	.90	.94	.90	.88
	Q3	.81	.87	.75	.79
	Q4	.94	.92	.92	.94
	Q5	.92	.96	.94	.94
Código faltante	Q1	.92	.96	.92	.90
	Q2	.98	.98	.96	.98
	Q3	.85	.88	.87	.87
	Q4	.92	.98	.92	.94
	Q5	.85	.88	.85	.87
Ordenamiento de pasos	Q1	.88	.94	.88	.88
	Q2	.94	.98	.96	.96
	Q3	.65	.85	.73	.73
	Q4	.88	.92	.90	.90
	Q5	.75	.88	.81	.81

Nota: V = coeficiente V de Aiken. Los intervalos de confianza al 95 % se calcularon para cada estimación.
Fuente: elaboración propia.

Tabla 3. Retención de ítems por forma y nivel de experiencia algorítmica

	IND-6	ISC-1	LSCA-3	LSCA-5
Forma de evaluación	(Nivel 1)	(Nivel 2)	(Nivel 3)	(Nivel 4)
Lectura de códigos	2	2	3	2
Resultados esperados	0	1	1	0
Código faltante	4	4	2	0
Ordenamiento de pasos	3	0	3	4

Fuente: elaboración propia.

faltante evidenció una disminución progresiva conforme aumentó la experiencia, pasando de cuatro ítems en los niveles iniciales a ninguno en el nivel más avanzado, lo que sugiere una posible reducción en su capacidad discriminativa en estudiantes con mayor dominio. En contraste, Ordenamiento de pasos destacó por su mayor consistencia en niveles superiores, alcanzando su máxima retención en el nivel 4, aunque con cierta variabilidad intermedia.

En conjunto, estos resultados ponen de manifiesto diferencias en el funcionamiento de las

distintas formas, lo que justifica su análisis diferenciado. Los indicadores específicos por ítem (IF, D y r_pb) se presentan en el anexo.

Para complementar el análisis psicométrico, se examinó el desempeño por nivel mediante la media del índice de facilidad (IF) en cada forma de evaluar. En Lectura de códigos se observó un incremento progresivo conforme aumenta la experiencia: de .58 (sin experiencia) y .64 (inicial) a .72 (intermedio) y .77 (intermedio–avanzado). En Resultados esperados, los valores fueron menores en los niveles sin experiencia (.42) e inicial (.50), con mejoras en intermedio (.49) e intermedio-avanzado (.59), más claras en este último. En Código faltante, el desempeño también aumentó con la experiencia, alcanzando .70 y .74 en niveles intermedio e intermedio-avanzado, mientras que los niveles sin experiencia e inicial se mantuvieron en .63. Finalmente, Ordenamiento de pasos mostró un patrón general de incremento (.52, .58, .64), aunque con un descenso en el nivel intermedio (.49), lo que sugiere cierta variabilidad.

En conjunto, los resultados sugieren que la dificultad y la calidad psicométrica de las diferentes formas no evolucionan de manera uniforme, lo que refuerza la necesidad de considerar el nivel de experiencia como un factor clave en el diseño y validación del instrumento.

Percepción versus *medición*

Al contrastar la facilidad percibida con el índice de facilidad (IF), se observan coincidencias parciales. En los niveles intermedio (LSCA-3) e intermedio-avanzado (LSCA-5), Lectura de códigos fue la más fácil, en concordancia con sus IF medios-altos (≈ .72 y .77); en el nivel inicial (IND-6) también fue la única percibida como fácil, con IF ≈ .58. En contraste, Código faltante presenta IF medios-altos (≈ .63-.74), pero es percibida como de las más difíciles, especialmente en IND-6 (53%) y en niveles intermedio e intermedio-avanzado (32%-37%). Ordenamiento de pasos fue la más difícil en todos los niveles, destacando ISC-1 (60%) y LSCA-5 (40%), consistente con IF moderados (≈ .49-.64). Finalmente, Resultados esperados no fue identificada consistentemente como fácil o difícil, pese a IF bajos-medios (≈ .42-.59), lo que sugiere menor claridad en su percepción.

Análisis de preguntas abiertas

Para la pregunta 1, se identificaron cuatro categorías principales: emociones positivas y de bienestar, confusión inicial o parcial, esfuerzo cognitivo y pensamiento analítico, y tensión o estrés (ver tabla 4). Los porcentajes corresponden a la proporción de estudiantes que mencionaron cada categoría, considerando que un mismo participante pudo reportar más de una emoción. En la selección de la forma más fácil y más difícil (tabla 5), se observa que una misma puede ser percibida como ambas, especialmente en Código faltante y Ordenamiento de pasos, lo que evidencia una percepción de dificultad no homogénea. Además, se analizaron de manera global las razones de elegir aquella percibida como más fácil. En Código Faltante, la facilidad se atribuyó a que "solo había que encontrar lo que faltaba" (37.5%) y a que "se entendía rápido y no era complicado" (29.2%). En Resultados esperados, predominó la idea de que "era más fácil de entender" y que "la lectura de códigos fue clara" (50%), junto con que "solo se seguían las indicaciones" (30%). En Ordenamiento de Pasos, se destacó que "ya se tenían los pasos y solo había que ordenarlos" y que "era más fácil ordenar que averiguar" (45%), así como "seguir un proceso" (35%). Finalmente, en Lectura de Códigos, la facilidad se explicó porque "ya se tenía una estructura definida" y "era fácil leer el código y saber qué iba a pasar" (50%), además de la posibilidad de "predecir" el comportamiento (30%) y la claridad de las indicaciones o apoyos previos (20%).

De manera análoga, para la percibida como más difícil: En Código faltante, la dificultad se asoció a que "había muchas posibilidades" y "las opciones resultaban confusas" (45%), así como a

la necesidad de "hacerlo paso a paso" (30%). En Resultados esperados, predominó la idea de que "me revolví con las instrucciones" y que "todo venía revuelto" (45%), junto con la exigencia de "buscar la ruta correcta" (30%). En Ordenamiento de pasos, la dificultad se atribuyó a la confusión inicial ("me confundía mucho") (50%) y a que "era más laborioso" (30%). Finalmente, en Lectura de códigos, se señaló que "no se tenía una estructura ordenada" (45%), junto con dificultades para "seguir la secuencia" (30%) y la percepción de múltiples soluciones posibles (25%). Las respuestas textuales matizan los resultados psicométricos al mostrar que la dificultad percibida no siempre coincide con el índice de facilidad (IF); por ejemplo, Código faltante presenta buen desempeño, pero se percibe como cognitivamente más demandante.

Finalmente, en las propuestas de mejora (ver tabla 6), varios estudiantes no sugirieron cambios y, entre quienes lo hicieron, destacaron la inclusión de más ejemplos, mayor claridad en las instrucciones y el uso de apoyos visuales, mientras que en menor medida se propuso ajustar la dificultad o hacer la prueba más dinámica o breve.

DISCUSIÓN

El presente estudio analizó el funcionamiento psicométrico y cognitivo de distintas formas de evaluación del pensamiento algorítmico en estudiantes

Tabla 4. Categorías emocionales reportadas por los estudiantes ante la pregunta: "¿Cómo te sentiste?"

	Grupos (%)				
Categoría	**IND-6**	**ISC-1**	**LSCA-3**	**LSCA-5**	**Promedio**
Emociones positivas y de bienestar	61	69	63	66	64.8
Confusión inicial o parcial	44	31	42	26	35.8
Esfuerzo cognitivo y pensamiento analítico	39	28	21	23	27.8
Tensión, nerviosismo o estrés	11	14	26	17	17.0

Fuente: elaboración propia.

Tabla 5. Preguntas 2 y 3: "¿Cuál fue la forma más fácil y más difícil?"

Forma de evaluación	Grupos (%)				Promedio
	ISC-1	IND-6	LSCA-3	LSCA-5	
Más Fácil					
Código faltante	28	28			28.0
Resultados esperados	28				28.0
Lectura de códigos		33	40	46	39.7
Ordenamiento de pasos			26	27	26.5
Más Difícil					
Código faltante	20	53	32	37	35.5
Resultados esperados					
Lectura de códigos					
Ordenamiento de pasos	60	29	39	40	42.0

Fuente: elaboración propia.

Los resultados respaldan una concepción progresiva de la evaluación del pensamiento algorítmico, en la cual las formas deben alinearse con el nivel de desarrollo del estudiante, evitando enfoques de evaluación únicos

universitarios con diferentes niveles de experiencia. Debido a su carácter de pilotaje, los resultados deben interpretarse como evidencia preliminar orientada al refinamiento del instrumento. En conjunto, se encontró un funcionamiento diferencial según el nivel de habilidad algorítmica: Lectura de códigos y Resultados esperados fueron más adecuadas en niveles intermedios e intermedios-avanzados, pero con limitaciones en niveles iniciales; Código faltante resultó más funcional en niveles iniciales e inicial-intermedios; y Ordenamiento de pasos presentó mayor estabilidad relativa, aunque con sensibilidad en etapas tempranas. Estos hallazgos evidencian que la adecuación de las formas de evaluar depende del nivel formativo del estudiante.

El desempeño no siguió un patrón estrictamente ascendente, lo que puede explicarse por la interacción entre factores cognitivos y psicométricos. Las formas implican demandas distintas –como simulación secuencial, reconocimiento estructural o planificación–, por lo que la experiencia no siempre se traduce en ventajas proporcionales (Robins *et al.*, 2003). Además, los niveles intermedios pueden representar fases de transición cognitiva (Lister *et al.*, 2004), mientras que la sensibilidad diferencial de los reactivos genera efectos de techo o piso (Crocker & Algina, 1986; Ebel & Frisbie, 1991). En este contexto, se identificaron tres patrones psicométricos: formas robustas transversalmente (Ordenamiento de pasos), sensibles al nivel (Lectura de códigos y Código faltante) y dependientes del contexto (Resultados esperados), en línea con los principios clásicos de análisis de reactivos (Crocker & Algina, 1986; Ebel & Frisbie, 1991). Asimismo, la retención de ítems mostró un comportamiento no lineal entre niveles, lo que sugiere que el desempeño depende también de procesos de reorganización cognitiva y variabilidad en las estrategias de resolución (Haladyna, 2016).

El contraste entre dificultad percibida y desempeño evidenció tanto convergencias como divergencias: Lectura de códigos fue percibida como accesible en niveles iniciales, mientras que Ordenamiento de pasos se identificó como más demandante; en cambio, Código faltante y Resultados esperados presentaron niveles de facilidad relativamente altos pese a ser percibidas como complejas, lo que indica que la dificultad subjetiva depende más de los procesos cognitivos implicados que del porcentaje de aciertos. En conjunto, estos resultados se explican por la interacción entre las demandas cognitivas de cada forma de evaluar y el desarrollo de los esquemas algorítmicos, cuya consolidación no es lineal. Esto se refleja también en la experiencia del estudiante, caracterizada por una activación cognitiva positiva, con confusión inicial transitoria y baja presencia de estrés, lo que sugiere una carga cognitiva adecuada y un uso formativo del instrumento.

En términos aplicados, los resultados respaldan una concepción progresiva de la evaluación del pensamiento algorítmico, en la cual las formas deben alinearse con el nivel de desarrollo del estudiante, evitando enfoques de evaluación únicos en contextos heterogéneos y favoreciendo el uso de instrumentos diferenciados o adaptativos. Desde una perspectiva pedagógica, los hallazgos apoyan enfoques escalonados, en los que distintas formas de evaluar, contribuyen al

desarrollo progresivo de habilidades algorítmicas y fortalecen la retroalimentación formativa. Al contrastar estos resultados con instrumentos previos, se observa que estos tienden a emplear un número limitado de formas o centrarse en una sola: el instrumento ATTA (Lafuente *et al*., 2022) integra Resultados esperados y Ordenamiento de pasos sin análisis psicométrico por cada forma; González (2014) se limita a Resultados esperados; Vasconcelos-Santillan (2008) incluye Lectura de códigos, Resultados esperados y Ordenamiento de pasos, pero con distribución desigual y análisis globales; y Navas (2022) privilegia ejercicios cualitativos centrados en Resultados esperados.

Esta tendencia también se observa en estudios realizados en México, donde la evaluación del pensamiento algorítmico se ha desarrollado principalmente en el marco de intervenciones tecno pedagógicas en universitarios (Velasco-Ramírez y Otero-Escobar, 2023; González, 2014), o mediante adaptaciones instrumentales con énfasis en la validez y consistencia interna, pero con limitaciones en la cobertura de formas de evaluación, como la ausencia de tareas de completado de algoritmos (Velasco y Sánchez, 2021). En conjunto, estos antecedentes muestran que es poco frecuente encontrar instrumentos usando varias formas con análisis psicométrico detallado según el nivel de experiencia.

CONCLUSIONES

El estudio aporta evidencia preliminar sobre el funcionamiento psicométrico de un instrumento con varias formas para evaluar el pensamiento algorítmico en estudiantes universitarios con distintos niveles de experiencia, mostrando que el desempeño no es homogéneo, sino dependiente del nivel formativo y de las demandas cognitivas específicas de cada una. Los resultados evidencian patrones diferenciados de funcionamiento, desde aquellas relativamente estables hasta otras sensibles al nivel o al contexto, así como un comportamiento no lineal del desempeño, lo que sugiere que el desarrollo del pensamiento algorítmico implica procesos de transición y reorganización cognitiva más que una progresión estrictamente ascendente.

Asimismo, el desempeño no sigue una progresión lineal, sino que presenta variaciones entre niveles, lo que sugiere la presencia de fases de transición en el desarrollo del pensamiento algorítmico, asociadas a procesos de reorganización cognitiva. En este marco, las discrepancias entre la dificultad percibida y el desempeño observado indican que la complejidad de las tareas no se explica únicamente por el porcentaje de aciertos, sino por los procesos cognitivos que activa cada forma, tales como la simulación secuencial, el reconocimiento estructural o la planificación.

No obstante, los resultados deben interpretarse considerando las limitaciones del estudio, particularmente su carácter de pilotaje, el tamaño reducido y no probabilístico de la muestra, y la posible calibración aún imperfecta de los reactivos, lo que restringe la generalización de los hallazgos y la estabilidad de algunos indicadores psicométricos. En este sentido, cobra especial relevancia el análisis de los reactivos clasificados como “por revisar”, cuyo desempeño limítrofe no

El desempeño no sigue una progresión lineal, sino que presenta variaciones entre niveles, lo que sugiere la presencia de fases de transición en el desarrollo del pensamiento algorítmico, asociadas a procesos de reorganización cognitiva

implica fallas estructurales, sino sensibilidad a condiciones como el nivel de habilidad del grupo o la formulación del ítem; por ello, ajustes puntuales en redacción, claridad de instrucciones o nivel de dificultad pueden optimizar su funcionamiento.

Como línea de trabajo futuro, se recomienda ampliar la muestra y su diversidad, incorporar análisis psicométricos más robustos, como modelos de teoría de respuesta al ítem, así como explorar el desarrollo de versiones adaptativas e implementaciones en entornos digitales interactivos que permitan evaluar de manera más precisa y dinámica los procesos cognitivos implicados en el pensamiento algorítmico.

REFERENCIAS

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taking motivation. *Journal of Experimental Psychology: Applied*, *15*(2), 163-181. https://doi.org/10.1037/a0015719

Adorni, G., & Piatti, A. (2025). Designing the virtual CAT: A digital tool for algorithmic thinking assessment in compulsory education. *International Journal of Child-Computer Interaction*, *45,* 10076. https://doi.org/1.1016/j.ijcci.2025.100760

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Waveland Press.

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214-221. https://doi.org/10.1016/j.sapharm.2018.03.066

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is Involved and what is the role of the computer science education community? *ACM Inroads*, *2*(1), 48-54. https://doi.org/1.1145/1929887.1929905

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, *2*, 8-14. https://doi.org/10.1016/j.npls.2016.01.001

Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking* [Conferencia]. Annual Meeting of the American Educational Research Association, Vancouver, BC, Canada. https://scratched.gse.harvard.edu/ct/files/AERA2012.pdf

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Holt, Rinehart and Winston.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, *119*(2), 166.e7-166.e16.

Denny, P., Luxton-Reilly, A., & Simon, B. (2008). *Evaluating a new exam question: Parsons problems* [Conferencia]. ICER'o8 International Workshop on Computing Education Research, Sidney, Australia. https://doi.org/10.1145/1404520.1404532

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement*. Prentice Hall.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107-115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine*, *7*(3), 93-99. https://doi.org/10.1016/j.afjem.2017.08.001

González Villarreal, R. J. (2014). *Efectos del aprendizaje basado en juegos en el pensamiento algorítmico de aprendices de programación* [tesis de maestría]. Repositorio Institucional, Universidad Autónoma de Aguascalientes. http://bdigital.dgse.uaa.mx:8080/xmlui/bitstream/handle/11317/1090/389736.pdf?sequence=1&isAllowed=y

Gómez Rodríguez, H., González Fernández, M. O., y Aceves Aldrete, C. E. (2025). Impacto del pensamiento computacional desde una mirada interdisciplinaria en la educación superior. *RISTI: Revista Ibérica de Sistemas e Tecnologias de Informação*, (57), 39-54. https://dialnet.unirioja.es/servlet/articulo?codigo=10386299

Guitert, M., y Romeu, T. (2022). *Competencias "Digitales" del siglo XXI.* Universitat Oberta de Catalunya. https://www.studocu.com/pe/document/universidad-tecnologica-del-peru/educacion/competencias-digitales/67679719

Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied measurement in education*, *2*(1), 37-50. https://doi.org/10.1207/s15324818ame0201_3

Haladyna, T. M. (2016). *Developing and validating multiple-choice test items*. Routledge. https://doi.org/1.4324/9781315819798

Méndez Hernández, S. O., y Bermúdez, J. F. (2023). El pensamiento computacional como competencia para el siglo XXI. *Ciencia Latina Revista Científica Multidisciplinar*, *7*(4), Article 4. https://doi.org/1.37811/cl_rcm.v7i4.7044

Lafuente Martínez, M., Lévêque, O., Benítez, I., Hardebolle, C., & Zufferey, J. D. (2022). Assessing computational thinking: Development and validation of the algorithmic thinking test for adults. *Journal of Educational Computing Research*, *60*(6), 1436-1463. https://doi.org/1.1177/07356331211057819

Lister, R., Adams, E. S., Fitzgerald, S., Fone, W., Hamer, J., Lindholm, M., McCartney, R., Moström, J. E., Sanders, K., Seppälä, O., Simon, B., & Thomas, L. (2004). A multinational study of reading and tracing skills in novice programmers. In *Proceedings of the 6th* Australasian Conference on Computing Education (pp. 119-150). https://doi.org/10.1145/1041624.1041673

Merino Soto, C. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales de Psicología*. https://www.um.es/analesps/v25/v25_1/19-25_1.pdf

Montero Sánchez, N. (2025). Pensamiento algorítmico en la enseñanza de matemáticas del nivel secundario: retos, estrategias y oportunidades. *Sciencevolution*, *4*(3), 47-60. https://doi.org/10.61325/ser.v4i3.206

Muñiz Fernández, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo: Revista del Colegio Oficial de Psicólogos*, *31*(1), 57-66. https://dialnet.unirioja.es/servlet/articulo?codigo=3150824

Navas-López, E. A. (2022). Una caracterización del desarrollo del pensamiento algorítmico de los estudiantes de las carreras de Licenciatura en Matemática y Licenciatura en Estadística de la sede central de la Universidad de El Salvador en el período 2018-202. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, (33), e15-e15. https://doi.org/1.24215/18509959.33.e15

Navas López, E. A. (2024). Relaciones entre la matemática, el pensamiento algorítmico y el pensamiento computacional. *IE Revista de Investigación Educativa de la REDIECH*, *15*, e1929-e1929. https://doi.org/1.33010/ie_rie_rediech.v15i.1929

Papert, S. (1980). Computers and Computer Cultures. En *Mindstorms Children, Computers, and Powerful Ideas*. Basic Books. https://rafaelleviblog.wordpress.com/wp-content/uploads/2016/11/mindstorms-chap1.pdf

Parsons, D., & Haden, P. (2006). *Parsons programming puzzles: A fun and effective learning tool for first programming courses* [Conferencia]. ACE' 06 8th Australasian Conference on Computing Education, Darlinghurst, Australia. https://dl.acm.org/doi/10.5555/1151869.1151890

Pinzón Pérez, D. F., y González Palacio, E. V. (2022). Incidencia de las habilidades de pensamiento algorítmico en las habilidades de resolución de problemas: Una propuesta didáctica en el contexto de la educación básica secundaria. *Estudios Pedagógicos (Valdivia)*, *48*(2), 415-433. https://doi.org/1.4067/S0718-07052022000200415

Pinzón Pérez, D. F., Román González, M., y González Palacio, E. V. (2023). El pensamiento algorítmico como estrategia didáctica para el desarrollo de habilidades de resolución de problemas en el contexto de la educación básica secundaria. *Revista de Educación a Distancia (RED)*, *23*(73). https://doi.org/1.6018/red.542111

Posso Domínguez, M. E., y Murcia Londoño, E. (2022). *Las "actividades desconectadas" y el desarrollo del pensamiento algorítmico*. Universidad Católica de Pereira. http://hdl.handle.net/10785/9635

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, *13*(2), 137-172. https://doi.org/10.1076/csed.13.2.137.14200

Rodríguez-Abitia, G., Ramírez-Montoya, M. S., López-Caudana, E. O., y Romero-Rodríguez, J. M. (2021). Factores para el desarrollo del pensamiento computacional en estudiantes de pregrado. *Campus Virtuales*, *10*(2), 153-164. http://uajournals.com/ojs/index.php/campusvirtuales/article/view/893

Rojas-López, A., y García-Peñalvo, F. J. (2020). Evaluación del pensamiento computacional para el aprendizaje de programación de computadoras en educación superior. *Revista de Educación a Distancia (RED)*, *20*(63). https://doi.org/1.6018/red.409991

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation*, *55*, 37-76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Vasconcelos-Santillan, J. (2008). *Development of a Web-Based Surveying Instrument to Identify Problem-Solving Abilities Related to Effective Instruction in Computer Programming* [tesis doctoral]. ProQuest LLC, Johns Hopkins University. https://www.proquest.com/openview/fe9b6f2d64e9202adcce8ce72ee53310/1?pq-origsite=gscholar&cbl=18750

Velasco-Ramírez, M. L., y Otero-Escobar, A. (2023). El impacto del uso de Scratch para favorecer el pensamiento algorítmico en estudiantes de educación superior. *Revista Innova Educación*, *5*(4), 105-12. https://doi.org/1.35622/j.rie.2023.05v.006

Velasco Ramírez, M. L., y Sánchez Hernández, G. L. (2021). Análisis de la competencia de resolución de problemas algorítmicos en

estudiantes de educación superior. En *Pensamiento Computacional en México* (pp. 53-71). Academia Mexicana de Computación, A. C. https://amexcomp.mx/media/publicaciones/pensamiento-comp-en-mexico-2021.pdf

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33-35. https://doi.org/10.1145/1118178.111821

Wing, J. M. (2008). Computational thinking and thinking about computing. *ResearchGate*. https://doi.org/1.1098/rsta.2008.0118

Wise, S. L. (2017). Rapid-guessing behavior: Its identification, interpretation, and implications. *Educational Measurement: Issues and Practice*, *36*(4), 52-61. https://doi.org/10.1111/emip.12165

Este artículo es de acceso abierto. Los usuarios pueden leer, descargar, distribuir, imprimir y enlazar al texto completo, siempre y cuando sea sin fines de lucro y se cite la fuente.

CÓMO CITAR ESTE ARTÍCULO:

Esquivel Gámez, I., Guerrero Posadas, M., y Olivares Ruiz, N. A. (2026). Evaluación del pensamiento algorítmico: análisis psicométrico de diferentes formas de evaluación en un estudio piloto. *Apertura*, *18*(2), 58-73. http://doi.org/10.32870/Ap.v18n2.2858

ANEXO. ÍNDICES PSICOMÉTRICOS DE LOS ÍTEMS POR MODALIDAD, GRUPO Y NIVEL DE EXPERIENCIA ALGORÍTMICA

	Institución 2								Institución 1							
Grupo	IND-6 (Nivel 1)				ISC-1 (Nivel 2)				LSCA-3 (Nivel 3)				LSCA-5 (Nivel 4)			
Modalidad-Ítem	IF	D	r_bp	St.	IF	D	r_bp	St.	IF	D	r_bp	St.	IF	D	r_bp	St.
Lectura de códigos																
Q1	.85	.32	.11	b	.89	.18	.31		.95	.11	.33		.93	.23	.27	
Q2	.63	.66	.3	a	.74	.41	.35	a	.72	.58	.22	a	.8	.38	-.05	
Q3	.78	.48	.06	b	.59	.33	.11	b	.65	.68	.43	a	.58	.77	.2	a
Q4	.41	.59	.3	a	.48	.66	.29	a	.56	.63	-.07		.73	.54	.13	b
Q5	.22	.35	.18		.48	.35	-.01		.7	.58	.35	a	.8	.54	.23	a
Resultados esperados																
Q1	.15	.18	-.11		.3	.38	-.14		.3	.15	-.04		.6	.64	.01	b
Q2	.59	.43	-.21		.63	.23	-.35		.67	.56	.18	b	.6	.64	.07	b
Q3	.15	.1	-.22		.19	.18	-.05		.21	.31	.22	b	.23	.24	-.13	
Q4	.52	.78	-.03		.63	.53	.17	a	.58	.43	.16	b	.7	.57	.17	b
Q5	.7	.63	.03	b	.74	.28	.02	b	.67	.46	.24	a	.8	.35	.13	b
Código ffaltante																
Q1	.42	.3	.01		.58	.88	.36	a	.63	.3	-.17		.84	.38	-.07	
Q2	.68	.8	.42	a	.81	.5	.26	a	.8	.42	.3	a	.94	.15	-.07	
Q3	.53	.68	.26	a	.27	.54	.24	a	.46	.47	.16	b	.38	.37	-.01	
Q4	.79	.6	.28	a	.73	.26	0		.85	.32	.28	b	.78	.28	-.04	
Q5	.74	.6	.34	a	.77	.63	.2	a	.76	.43	.35	a	.78	.28	-.04	
Ordenamiento de pasos																
Q1	.63	.67	.35	a	.52	.18	-.1		.41	.22	-.18		.72	.6	.34	a
Q2	1	0			.8	.16	.3		.73	.35	.25	a	.94	.2	.26	
Q3	.37	.69	.19	a	.36	.08	-.24		.37	.62	.24	a	.53	.56	.24	a
Q4	.37	.8	.33	a	.68	.12	-.01		.49	.46	.19	b	.53	.86	.44	a
Q5	.21	.29	-.06		.56	.28	.19	b	.44	.47	.24	a	.5	.69	.31	a

IF = Índice de Facilidad. D = Índice de Discriminación. r_bp = Coeficiente de Discriminación biserial St. = Status de evaluación (a=Conservar, b=Revisar)
Fuente: elaboración propia.