Journal of Research in Educational Systems

Volume 19, Issue 70, 2025
Pp. 17-32

Print ISSN: 2383-1324
Online ISSN: 2783-2341

Homepage: www.jiera.ir

The Landscape of Computational Thinking Assessment in Primary Education: A Critical Review of Tools and Research Trends

Mona Batoei Avarzaman [1], **Maryam Mohsenpour** [2✉], and **Abolfazl Rafiepour** [3]

1. Ph.D. Student, Department of Educational Psychology, Faculty of Education and Psychology, Alzahra University, Tehran, Iran. E-mail: ghiasi@atu.ac.ir
2. *Corresponding Author*, Associate Professor, Department of Educational Psychology, Faculty of Education and Psychology, Alzahra University, Tehran, Iran. E-mail: ghiasi@atu.ac.ir
3. Associate Professor of Mathematics Education, Department of Mathematics Education, Faculty of Mathematics and Computer- Shahid Bahonar University of Kerman & Mahani Math Center, Afzalipour Research Institute, Shahid Bahonar University of Kerman, Kerman, Iran. E-mail: ghiasi@atu.ac.ir

ABSTRACT

Objective: Computational thinking is a key competency in primary education; however, its valid measurement faces serious conceptual and psychometric challenges. Beyond a mere classification of tools, this study presents a diagnostic analysis of the field through a dual approach: first, mapping the scientific landscape via scientometric analysis; and second, critically evaluating existing assessment tools.

Method: In this systematic critical review (PRISMA 2020), 45 eligible articles were identified by searching Scopus and Google Scholar databases (2006-2025). Analyses were conducted in two phases: 1) Scientometric analysis to map the conceptual structure of the field using R and VOSviewer software, and 2) Qualitative critical analysis of 8 selected key tools regarding their theoretical foundations and psychometric properties.

Results: The scientometric analysis revealed a rapidly growing but unbalanced field. The keyword co-occurrence map indicated a strong focus on "educational applications" and "programming tools," contrasted by a significant absence of assessment concepts such as "validity" and "fairness." This structural gap is directly reflected in the tools, revealing three main limitations: 1) Psychometric weakness due to reliance on traditional metrics (Cronbach's alpha) and neglect of modern standards (such as Omega coefficient and DIF analysis); 2) Conceptual limitations by focusing on algorithmic knowledge while ignoring process skills (debugging) and attitudinal aspects (perseverance); and 3) An instrumental gap for upper primary students.

Conclusions: Based on the findings, no single tool is sufficient for the comprehensive assessment of computational thinking. The maturity of this field requires adopting modern psychometric standards such as the Omega reliability coefficient, designing and localizing tests for primary students in Iran, and creating hybrid assessment frameworks that integrate standardized tests with performance-based assessments. This integrated approach will lead to more accurate skill diagnosis and the design of more effective educational programs.

Article Info:

Article Type:
Research Article

Article history:
Received July 04, 2025
Received in revised form August 23, 2025
Accepted September 01, 2025
Published Online September 27, 2025

***Keywords*:**
assessment tools,
computational thinking ,
primary students ,
critical review ,
psychometric properties

Cite this article: Batoei Avarzaman, M., Mohsenpour, M., & Rafiepour, A. (2025). The Landscape of Computational Thinking Assessment in Primary Education: A Critical Review of Tools and Research Trends. *Journal of Research in Educational Systems*, *19*(70), 17-32. https://doi.org/10.22034/jiera.2026.551273.3385

© The Author(s)
Publisher: Iranian Educational Research Association

Introduction

Rapid technological advancements have positioned Computational Thinking (CT) as a critical 21st-century competency. While Wing (2006) defined CT as a universal problem-solving skill, the field remains theoretically fragmented between "tool-centric" and "competency-centric" paradigms. This ambiguity complicates assessment, leading to a disconnect between existing tools and psychometric standards. Although various measurement approaches exist—ranging from standardized tests to code analysis in environments like Scratch—recent reviews (Bass et al., 2020; Ocampo et al., 2024) highlight a significant lack of validity evidence.

This gap is especially critical for primary school students (ages 6–12), where assessment faces unique challenges such as limited literacy and test anxiety. Relying on unvalidated tools in this "concrete operational" stage can compromise educational outcomes, particularly in the Iranian context where applied problem-solving is a known weakness. Consequently, this study aims to bridge this gap through a systematic critical review. Adopting a dual approach, it maps the scientific landscape via scientometric analysis and critically evaluates the psychometric properties of CT assessment tools to provide a diagnostic framework for future interventions.

Method

This study was conducted as a systematic critical review adhering to the PRISMA 2020 guidelines (Page et al., 2021). Employing a mixed-methods approach (quantitative and qualitative), the research aimed to map the scientific landscape of CT assessment and critically analyze tools designed for primary education. A systematic search was performed across Scopus and Google Scholar, supplemented by ERIC and IEEE Xplore, covering the period from 2006 to 2025. The search string utilized Boolean operators combining keywords such as "computational thinking," "assessment," "scale," and "instrument."

Following the screening of 674 initial records based on inclusion criteria (focus on primary students aged 6–12, quantitative reporting of tool construction, and English publication) and exclusion criteria (reviews, theses, and non-relevant populations), 45 eligible studies were identified. From this corpus, 8 key instruments were selected for in-depth analysis based on psychometric reporting quality and accessibility.

Data analysis proceeded in two phases. First, a scientometric analysis was conducted using the *bibliometrix* R package and VOSviewer to map publication trends and conceptual structures. This phase identified the most frequently used tools, directly informing the selection for the second phase. Second, a critical content analysis of the selected tools was performed using a researcher-made data extraction form focusing on target populations, components, and psychometric properties. To ensure rigor, the checklist's validity was confirmed by three experts, and inter-rater reliability was established by double-coding 20% of the articles.

Results

The findings are presented in two main sections: scientometric analysis and critical analysis of assessment tools.

Scientometric Analysis: Trends and Structure

The analysis reveals an exponential growth in publications starting from 2015, peaking in 2023, indicating the field's rapid expansion. Research is geographically concentrated in the US and China, followed by Spain and Taiwan, raising concerns about the cultural generalizability of tools for contexts like Iran. The co-occurrence map exposes a profound structural gap: the network is dominated by "educational applications" and "programming tools," while key psychometric concepts like "validity" are virtually absent. This suggests a prioritization of instructional design over rigorous assessment.

Critical Analysis of Assessment Tools

The micro-level analysis of 8 selected tools mirrors these macro-trends. While instruments like the Beginners' CT Test (BCTt)and TechCheck

demonstrate strong content validity and reliability (α > 0.80) for lower primary students (ages 5–9), they suffer from "ceiling effects" when applied to older students (ages 9–12). Conversely, tools like the CT Competency Test (cCTt) use advanced Item Response Theory (IRT) to ensure gender fairness but still lack difficult items for high-ability learners.

A recurring limitation across most tools (e.g., CTtLP and CTTC) is the exclusive reliance on multiple-choice formats. While efficient, this format restricts assessment to conceptual knowledge (sequences, loops), failing to capture process-oriented skills like debugging or attitudinal dimensions like perseverance. Furthermore, experimental tools like ACES show limited reliability (α=0.68) and lack comprehensive validity evidence. In summary, no single tool offers a complete solution; there is a persistent trade-off between psychometric rigor and the breadth of skills assessed.

Conclusions

The findings of this review reveal a structural imbalance in the field of Computational Thinking (CT) assessment. While the scientometric analysis highlights a research ecosystem dominated by educational applications, the critical analysis of tools exposes a significant gap in psychometric rigor. Most existing instruments rely on traditional metrics like Cronbach's alpha, largely neglecting modern standards such as McDonald's Omega and Differential Item Functioning (DIF) analysis (AERA, APA, & NCME, 2014). This technical shortfall aligns with a "conceptual imbalance," where assessments disproportionately focus on algorithmic concepts (e.g., loops) amenable to multiple-choice formats, at the expense of process-oriented skills (e.g., debugging) and attitudes.

Furthermore, the geographical concentration of research in the US and China poses a serious challenge for cultural validity in contexts like Iran. The lack of localized, standardized tools suggests that mere translation is insufficient. To address these systemic issues, this study proposes a three-phase roadmap for future research:

1) **Psychometric Upgrade:** Adopting modern item response theory (IRT) models and fairness analyses to ensure valid measurement;

2) **Holistic Assessment:** Moving towards hybrid models that integrate standardized tests with performance-based assessments (e.g., project portfolios) to capture the full spectrum of CT competence; and

3) **Intelligent Localization:** Developing context-sensitive instruments tailored to national curricula and cultural nuances. Implementing this roadmap is essential for empowering educators and policymakers to make evidence-based decisions, shifting the focus from the proliferation of new tools to the rigorous validation and standardization of existing ones.

Ethical Considerations

This study was conducted as part of a Ph.D. dissertation approved by the Ethics Committee of Alzahra University (Ethical code: IR.ALZAHRA.REC.1404.081). The authors strictly adhered to ethical principles, avoiding data fabrication, falsification, plagiarism, and research misconduct.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no conflict of interest.

پژوهش در نظام‌های آموزشی

دوره ۱۹، شماره ۷۰، ۱۴۰۴
ص ۱۷-۳۲

شاپا (چاپی): ۱۳۲۴-۲۳۸۳
شاپا (الکترونیکی): ۲۳۴۱-۲۷۸۳

Homepage: www.jiera.ir

چشم‌انداز سنجش تفکر محاسباتی در دورهٔ ابتدایی: مروری نقّادانه بر ابزارها و روندهای پژوهشی

مونا بطویی آورزمان [۱]، مریم محسن پور [۲]✉، و ابوالفضل رفیع پور [۳]

۱. دانشجوی دکتری تخصصی روان‌شناسی تربیتی، گروه روان‌شناسی تربیتی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه الزهرا، تهران، ایران. رایانامه: m.batoei@alzahra.ac.ir

۲. *نویسنده مسئول*، دانشیار، گروه روان‌شناسی تربیتی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه الزهرا، تهران، ایران. رایانامه: m.mohsenpour@alzahra.ac.ir

۳. دانشیار بخش آموزش ریاضی، دانشکده ریاضی و کامپیوتر، دانشگاه شهید باهنر کرمان و رییس گروه پژوهشی آموزش ریاضی، پژوهشکده ریاضی ماهانی، پژوهشگاه افضلی پور، دانشگاه شهید باهنر کرمان، کرمان، ایران. رایانامه: rafiepour@uk.ac.ir

چکیده

هدف: تفکر محاسباتی یک شایستگی کلیدی در آموزش ابتدایی است، اما سنجش معتبر آن با چالش‌های مفهومی و روان‌سنجی جدی مواجه است. این پژوهش با ارائهٔ یک تحلیل تشخیصی، فراتر از طبقه‌بندی صرف ابزارها رفته و با رویکردی دوگانه به این حوزه می‌پردازد: نخست، ترسیم نقشه علمی از طریق تحلیل علم‌سنجی؛ و دوم، ارزیابی نقّادانه ابزارهای سنجش موجود.

روش: در این مرور نقّادانه نظام‌مند (PRISMA, 2020)، با جستجو در پایگاه‌های Scopus و Google Scholar (۲۰۰۶–۲۰۲۵)، تعداد ۴۵ مقاله واجد شرایط شناسایی شد. تحلیل‌ها در دو مرحله انجام گرفت: ۱) تحلیل علم‌سنجی برای ترسیم ساختار مفهومی حوزه با استفاده از نرم‌افزارهای R و VOSviewer، و ۲) تحلیل نقّادانه کیفی ۸ ابزار کلیدی منتخب از منظر مبانی نظری و ویژگی‌های روان‌سنجی.

یافته‌ها: تحلیل علم‌سنجی، میدانی با رشد سریع اما نامتوازن را آشکار ساخت. نقشهٔ هم‌رخدادی کلیدواژه‌ها، تمرکز شدید بر «کاربردهای آموزشی» و «ابزارهای برنامه‌نویسی» را در مقابل غیبت معنادار مفاهیم سنجش مانند «روایی» و «انصاف» نشان داد. این شکاف ساختاری، مستقیماً در ابزارها منعکس شده و سه محدودیت اصلی را نمایان می‌کند: ۱) ضعف روان‌سنجی به دلیل اتکا به معیارهای سنتی (آلفای کرونباخ) و غفلت از استانداردهای مدرن (مانند شاخص امگا و تحلیل DIF)؛ ۲) محدودیت مفهومی با تمرکز بر دانش الگوریتمی و نادیده گرفتن مهارت‌های فرایندی (اشکال‌زدایی) و نگرشی (پشتکار)؛ و ۳) خلأ ابزاری برای دانش‌آموزان دورهٔ دوم ابتدایی.

نتیجه‌گیری: بر مبنای یافته‌های به‌دست آمده، هیچ ابزار واحدی برای ارزیابی جامع تفکر محاسباتی کافی نیست. بلوغ این حوزه مستلزم به‌کارگیری استانداردهای روان‌سنجی مدرن مانند شاخص پایایی امگا، طراحی و بومی‌سازی آزمون‌ها برای دانش‌آموزان ابتدایی در ایران و ایجاد چهارچوب‌های ارزیابی ترکیبی که آزمون‌های استاندارد را با سنجش‌های عملکردی ادغام می‌کنند، متمرکز شوند. این رویکرد یکپارچه به تشخیص دقیق‌تر مهارت‌ها و طراحی برنامه‌های آموزشی مؤثرتر منجر خواهد شد.

درباره مقاله

نوع مقاله:
مقاله پژوهشی

تاریخچه مقاله:
تاریخ دریافت: ۱۴۰۴/۰۴/۱۳
تاریخ بازنگری: ۱۴۰۴/۰۶/۰۱
تاریخ پذیرش: ۱۴۰۴/۰۶/۱۰
تاریخ انتشار: ۱۴۰۴/۰۷/۰۵

واژه‌های کلیدی:
ابزارهای سنجش،
تفکر محاسباتی،
دانش‌آموزان ابتدایی،
مرور نقّادانه،
ویژگی‌های روان‌سنجی

استناد به این مقاله: بطویی آورزمان، م.، محسن‌پور، م.، و رفیع‌پور، ا. (۱۴۰۴). چشم‌انداز سنجش تفکر محاسباتی در دورهٔ ابتدایی: مروری نقّادانه بر ابزارها و روندهای پژوهشی. *پژوهش در نظام‌های آموزشی*، ۱۹(۷۰)، ۱۷-۳۲.
https://doi.org/10.22034/jiera.2026.551273.3385

© نویسندگان

ناشر: انجمن پژوهش‌های آموزشی ایران

مقدمه

تحولات سریع فناوری و ظهور هوش مصنوعی[1]، نظام‌های آموزشی را ناگزیر به بازنگری در مهارت‌های مورد نیاز نسل آینده کرده است. یکی از مهم‌ترین پاسخ‌های آموزشی به این تغییرات، تمرکز بر توسعهٔ مهارت‌هایی است که مستقیماً با فناوری‌های دیجیتال در ارتباط‌اند. در همین راستا، تفکر محاسباتی[2] به عنوان یکی از مهارت‌های کلیدی قرن بیست و یکم، به‌طور روزافزون در کانون توجه پژوهشگران و سیاست‌گذاران آموزشی قرار گرفته است (Haseski et al., 2018). اهمیت این مهارت به‌ویژه در بافت آموزشی ایران دوچندان است؛ چراکه پژوهش‌های ملی و بین‌المللی نشان‌دهندهٔ ضعف دانش‌آموزان ایرانی در مهارت‌های حل مسئله کاربردی هستند (شایان و یافتیان، ۱۳۹۷؛ ضیاءنژاد و دیگران، ۱۴۰۱). از این رو، بسیاری بر لزوم ادغام آن در برنامه‌های درسی به عنوان یک شایستگی بنیادین برای تقویت یادگیری و فراشناخت تأکید دارند (Yadav et al., 2022; Grover & Pea, 2013).

تفکر محاسباتی، که ریشه در علوم کامپیوتر دارد، فراتر از مهارت فنی برنامه‌نویسی است. (Wing, 2006) آن را «فرایند حل مسئله، طراحی سیستم‌ها و درک رفتارهای انسانی با تکیه بر مفاهیم علوم کامپیوتر» تعریف کرد و بر ضرورت آن برای همگان تأکید ورزید (ص.۳۳). با وجود پذیرش گسترده تعریف وینگ، ادبیات پژوهشی همچنان با فقدان اجماع بر سر مؤلفه‌های دقیق این سازه مواجه است. تحلیل عمیق‌تر نشان می دهد که این تکثر، ناشی از تفاوت در اهداف آموزشی و پارادایم‌های حاکم بر حوزه‌های علوم کامپیوتر و علوم تربیتی است. از یک سو، رویکرد ابزار–محور (متأثر از علوم کامپیوتر) تفکر محاسباتی را مهارتی فنی برای «فرمول‌بندی مسائل جهت حل توسط عامل پردازشگر» می‌داند. در این پارادایم، مدل‌هایی همچون (Brennan & Resnick, 2012) قرار دارند که تمرکزشان بر خلق مصنوعات دیجیتال، اشکال‌زدایی کد و مدل‌سازی سیستم‌هاست. از سوی دیگر، رویکرد شایستگی–محور (متأثر از علوم تربیتی و شناختی) این سازه را به‌عنوان یک مهارت شناختی عام و فراتر از کُدنویسی می‌بینند. مدل‌هایی همچون (Selby & Woollard, 2013) در این دسته جای می‌گیرند که بر مؤلفه‌های ذهنی انتزاعی مانند تجزیه، انتزاع و تفکر منطقی تأکید دارند و هدفشان ادغام این مهارت در آموزش عمومی (مانند ریاضی و علوم) است.

این شکاف میان رویکرد کُدنویسی‌محور و رویکرد شناختی، مستقیماً به پیچیدگی و ناهمگونی در حوزهٔ سنجش دامن زده است؛ چراکه طراحان ابزار ناگزیرند پیش از هر اقدامی، موضع خود را در یکی از این دو پارادایم مشخص کنند. در میان این مدل‌ها، چهارچوب متداول و پرارجاع (Brennan & Resnick, 2012) با تأکید بر سه بُعد مفاهیم محاسباتی[3]، شیوه‌های محاسباتی[4] و دیدگاه‌های محاسباتی[5]، سهم بسزایی در عملیاتی‌سازی مفهوم تفکرمحاسباتی داشته است. با این حال، یک مرور نقّادانه نیازمند توجه به محدودیت‌های این مدل نیز هست. همان‌طور که (Zhang, & Nouri, 2019) اشاره کرده‌اند، چهارچوب (Brennan & Resnick, 2012) به دلیل وابستگی شدید به محیط بصری اسکرچ، در تبیین تفکر محاسباتی در بافت‌های بدون کامپیوتر[6] یا حل مسائل دنیای فیزیکی دچار محدودیت است. همچنین، سنجش دقیق و کمیِ بُعد دیدگاه‌های محاسباتی (مانند بیانگری و تعامل) همواره با ابهام و چالش‌های روش‌شناختی و روان‌سنجی روبرو بوده است (Hsu & Chen, 2025).

انعکاس این چالش‌های نظری در حوزهٔ عمل، منجر به شکل‌گیری طیف وسیعی از رویکردهای سنجش شده است. مرور ادبیات پژوهشی حاکی از آن است که تلاش‌ها برای عملیاتی‌سازی و سنجش تفکر محاسباتی در چهار جریان اصلی متبلور شده است. در جریان نخست، پژوهشگران با توسعه آزمون‌های استاندارد و مقیاس‌های کمّی (González et al., 2017; Li & Liu, 2021; Zapata-Caceres & Martín-Barroso, 2024) و بهره‌گیری از مدل‌سازی ساختاری در پژوهش‌های داخلی (رمضانی و شریفی، ۱۴۰۳) کوشیده‌اند تا ابزاری عینی و سنجش‌پذیر فراهم آورند. به موازات این تلاش‌های کمّی، رویکرد دوم بر استفاده از محیط‌های برنامه‌نویسی بصری نظیر اسکرچ تمرکز داشته است. یافته‌های حاصل از این رویکرد در مطالعات داخلی (مؤمنی و نژادحسینیان، ۱۴۰۱؛ رزبان و همکاران، ۱۳۹۹) و خارجی (Fagerlund et al., 2021) نشان می‌دهد که اگرچه تحلیل پروژه‌های اسکرچ تصویری از مهارت‌های عملی دانش‌آموزان ارائه می‌دهد، اما در پوشش ابعاد درونی‌تر مانند نگرش‌ها و فرایندهای حل مسئله با محدودیت مواجه است.

---

1 artificial intelligence
2 computational thinking
3 computational concepts
4 computational practices
5 computational perspectives
6 unplugged

از سوی دیگر، رویکرد سوم با بهره‌گیری از رباتیک آموزشی (Stewart et al.,2021) تلاش کرده است تا سنجش را در بافتی ملموس و فیزیکی پیاده‌سازی کند؛ هرچند دقتِ سنجش مهارت‌های شناختی در این روش گاهی تحت‌الشعاع جذابیت ابزار قرار می‌گیرد. در نهایت، رویکردهای کیفی (مانند پژوهش روحی و دیگران، ۱۴۰۲) اگرچه سهم بسزایی در شناسایی ابعاد سازه داشته‌اند، اما فاقد قابلیت تبدیل شدن به ابزارهای استاندارد کلاسی هستند.

با وجود این تکثر در روش‌های سنجش، مرورهای نظام‌مند اخیر بر وجود یک شکاف روش‌شناختی جدی تأکید کرده‌اند. برخلاف تصور رایج دربارهٔ غنای ابزاری این حوزه، (Tang et al., 2020) در مرور خود صراحتاً بیان می‌کنند که اکثریت مطالعات بررسی‌شده، گزارش ناقصی از ویژگی های روان‌سنجی (روایی و پایایی) ارائه داده‌اند و کیفیت فنی ابزارها همچنان مبهم است. همسو با این یافته، (Ocampo et al.,2024) نیز نتیجه گرفتند که ابزارهای موجود فاقد شواهد کافی برای روایی سازه در گروه‌های سنی پایین می‌باشند. به عبارت دیگر، ادبیات موجود مملو از پاسخ به پرسش «چه ابزارهایی وجود دارند؟» است، اما در پاسخ به پرسش «این ابزارها چقدر معتبر هستند؟» با فقر شواهد مواجه است.

این پیچیدگی‌های نظری و عملی در طراحی ابزار، به‌ویژه در دورهٔ ابتدایی (سنین ۶ تا ۱۲ سال) برجسته‌تر است. از منظر شناختی، طبق نظریهٔ پیاژه، این مقطع مصادف با مرحله «عملیات عینی» است؛ دورانی حساس که در آن پایه‌های منطق و درک قواعد سیستماتیک در ذهن کودک شکل می‌گیرد و گذار از تفکر پیش‌عملیاتی به عملیاتی رخ می‌دهد. بنابراین، سنجش تفکر محاسباتی در این سنین نیازمند ابزارهایی است که مفاهیم انتزاعی را به صورت ملموس و عینی ارائه دهند. از منظر آموزشی نیز، پژوهش‌ها نشان داده‌اند که شکل‌گیری عادات صحیح فکری و حل مسئله در این دوره، نقش بنیادینی در موفقیت‌های تحصیلی آینده در حوزه‌های STEM ایفا می‌کند (Yadav et al., 2022).

با این حال، سنجش دانش‌آموزان این ردهٔ سنی (به‌ویژه زیر ۱۰ سال) با چالش‌های منحصربه‌فردی روبروست. همان‌طور که (Zhang, & Nouri, 2019) اشاره کرده‌اند، محدودیت در مهارت‌های شناختی خاص، به ویژه مهارت خواندن کُد، استفاده از ابزارهای متنی را ناممکن می‌سازد. علاوه بر این، محدودیت‌های عملی همچون سطح سواد خواندن پایین، دامنهٔ توجه محدود و اضطراب آزمون، همگی عواملی هستند که روایی و پایایی ابزارهای معمول (مانند پرسشنامه‌های لیکرتی طولانی) را در این گروه سنی تهدید می‌کنند. از این رو، تمرکز این پژوهش بر بازهٔ سنی ۶ تا ۱۲ سال، تلاشی برای واکاوی ابزارهایی است که توانسته‌اند بر این موانع شناختی و رشدی غلبه کنند.

ضرورت و اهمیت این پژوهش از آنجا ناشی می‌شود که همان‌طور که (Tang et al., 2020) اشاره کرده‌اند، استفاده از ابزارهای معتبر در حدود نیمی از مطالعات تفکر محاسباتی نادیده گرفته شده است. این غفلت، به‌ویژه در دورهٔ ابتدایی که چالش‌های عملی سنجش (مانند بار شناختی و محدودیت‌های کلامی) اعتبار نتایج را تهدید می‌کنند، نگران‌کننده‌تر است. از همین رو، امروزه مبحث سنجش در خط مقدم پژوهش‌های تفکر محاسباتی قرار گرفته است (Tikva & Tambouris, 2021). بر همین اساس، هدف پژوهش حاضر، پر کردن این شکاف از طریق انجام یک مرور نقّادانه بر ابزارهای طراحی‌شده برای سنجش تفکر محاسباتی در دانش‌آموزان دورهٔ ابتدایی است. این پژوهش با تحلیل یکپارچه مؤلفه‌های مورد سنجش، ویژگی‌های روان‌سنجی و قابلیت‌های بومی‌سازی، نقاط قوت و ضعف هر یک از ابزارها را تحلیل کرده و تصویری شفاف از وضعیت فعلی و مسیرهای آتی در این حوزه، ارائه می‌دهد.

روش

پژوهش حاضر در قالب یک مرور نقّادانه[1] نظام‌مند و با پیروی دقیق از دستورالعمل گزارش‌دهی PRISMA 2020 (Page et al., 2021) انجام شد. این مطالعه با رویکردی ترکیبی (کمّی و کیفی)، دو هدف اصلی را دنبال کرد: نخست، ترسیم نقشهٔ علمی حوزهٔ سنجش تفکر محاسباتی از طریق تحلیل علم‌سنجی؛ و دوم، شناسایی و تحلیل عمیق ابزارهای سنجش ویژه دانش‌آموزان دورهٔ ابتدایی از منظر مبانی نظری، ویژگی‌های روان‌سنجی و تناسب کاربردی.

به منظور شناسایی جامع مطالعات مرتبط، جستجوی نظام‌مندی در پایگاه استنادی معتبر Scopus (به عنوان پایگاه مرجع اصلی) و موتور جستجوی Google Scholar (برای پوشش ادبیات خاکستری) در بازه زمانی ۲۰۰۶ (همزمان با معرفی رسمی مفهوم تفکر محاسباتی توسط وینگ) تا ۲۰۲۵ صورت گرفت. همچنین جهت اطمینان از عدم حذف مطالعات کلیدی، بررسی تکمیلی در پایگاه‌های تخصصی ERIC و IEEE Xplore

1. critical review

نیز انجام شد. رشته جستجوی[1] با استفاده از عملگرهای منطقی و ترکیبی از کلیدواژه‌های اصلی و مترادف‌های مرتبط با «تفکر محاسباتی» و «ابزارهای سنجش» از جمله «مقیاس»، «ابزار»، «سنجش» و «آزمون» طراحی گردید.

فرایند غربالگری و انتخاب مطالعات مطابق نمودار جریان پریسما (شکل ۱) پیگیری شد. با توجه به ماهیت الگوریتم جستجوی گوگل اسکالر، غربالگری در این پایگاه با پروتکل بررسی ۲۰۰ نتیجه نخست (مرتب‌سازی بر اساس میزان ارتباط) انجام گرفت. جهت مدیریت بهینه داده‌ها، کلیهٔ رکوردهای مستخرج در یک پایگاه داده مرجع تجمیع شدند و عملیات پالایش و حذف موارد تکراری طی یک فرایند دو مرحله‌ای (شناسایی سیستمی و بازبینی دستی) صورت پذیرفت. از مجموع ۶۷۴ پیشینه شناسایی‌شدهٔ اولیه، پس از غربالگری چندمرحله‌ای و اعمال معیارهای ورود (شامل تمرکز بر جامعهٔ دانش‌آموزان دورهٔ ابتدایی (پایه‌های اول تا ششم) با بازهٔ سنی ۶ تا ۱۲ سال، ارائهٔ گزارش کمّی از ساخت ابزار و انتشار به زبان انگلیسی) و خروج (حذف مقالات مروری، همایشی، پایان‌نامه‌ها، کتاب‌ها و فصول کتاب، مطالعات با جامعهٔ پژوهشی غیرمرتبط)، در نهایت ۴۵ پژوهش انتخاب شدند. از میان این مجموعه، ۸ ابزار شاخص نه صرفاً بر اساس بسامد تکرار، بلکه با معیارهای ترکیبی شامل کیفیت گزارش‌دهی روان‌سنجی و دسترسی‌پذیری جهت واکاوی عمیق برگزیده شدند.

فرایند استخراج و تحلیل داده‌ها در دو مرحلهٔ مجزا انجام پذیرفت. در گام نخست، به منظور تحلیل علم‌سنجی و ترسیم نقشه علمی حوزه، داده‌های کتاب‌شناختی با استفاده از بستهٔ نرم افزاری bibliometrix در محیط برنامه نویسی R و نرم‌افزار VOSviewer تحلیل شدند. با هدف پاسخ به سه پرسش کلیدی انجام شد: ۱) روند رشد کمّی انتشارات در حوزه سنجش تفکر محاسباتی چگونه است؟ ۲) ساختار مفهومی و خوشه‌های موضوعی غالب کدام‌اند؟ و ۳) کلیدواژه‌ها و ابزارهای پرتکرار در ادبیات تحقیق کدام موارد هستند؟ نتایج این بخش صرفاً توصیفی نیست، بلکه به عنوان مبنای غربالگری برای گام دوم عمل می‌کند؛ بدین معنا که شناسایی پربسامدترین ابزارها در نقشهٔ علمی، مستقیماً فهرست ابزارهای منتخب برای تحلیل نقّادانه را تعیین کرده است.

در گام دوم، تحلیل محتوایی و نقادانهٔ ابزارهای منتخب بر اساس یک فُرم استخراج داده پژوهشگرساخته (شامل جامعهٔ هدف، مؤلفه‌ها و ویژگی های روان‌سنجی) صورت پذیرفت. به منظور تضمین کیفیت، روایی صوری و محتوایی این چک‌لیست توسط ۳ متخصص روان‌سنجی و علوم تربیتی تأیید گردید و جهت اطمینان از پایایی استخراج داده‌ها، ۲۰ درصد از مقالات به صورت تصادفی توسط پژوهشگر دوم کُدگذاری و توافق حاصل شد.

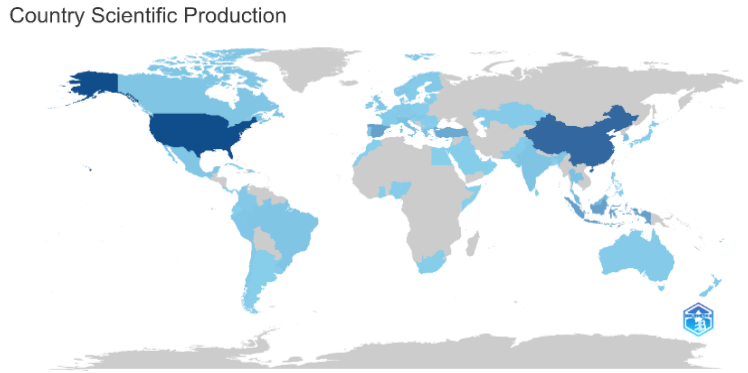
شکل ۱. نمودار جریان انتخاب پژوهش‌ها بر اساس چهارچوب پریسما ۲۰۲۰

1. query

یافته‌ها

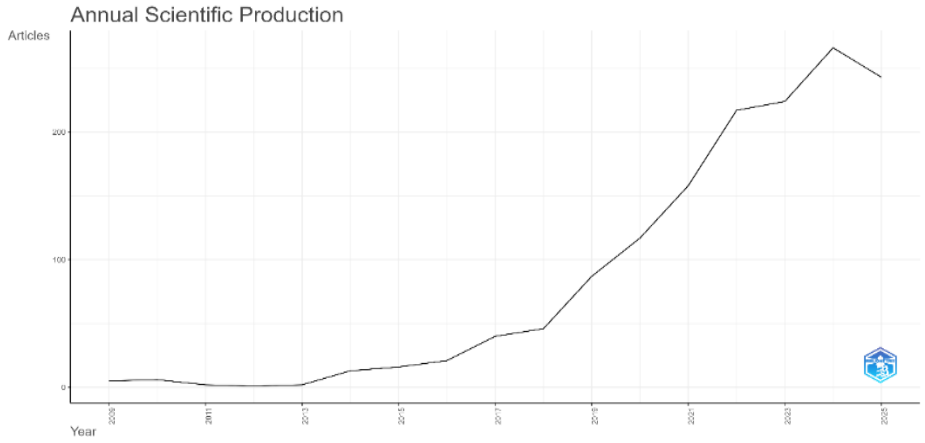
یافته‌های پژوهش در دو بخش اصلی ارائه می‌شود. بخش نخست به تحلیل علم‌سنجی حوزه اختصاص دارد و بخش دوم به تحلیل نقّادانه ابزارهای سنجش تفکر محاسباتی می‌پردازد.

۱. تحلیل علم‌سنجی: روندها، کانون‌ها و ساختار حوزهٔ پژوهش



شکل ۲. روند زمانی تولیدات علمی سالانه در حوزه سنجش تفکر محاسباتی (منبع: یافته‌های تحقیق)

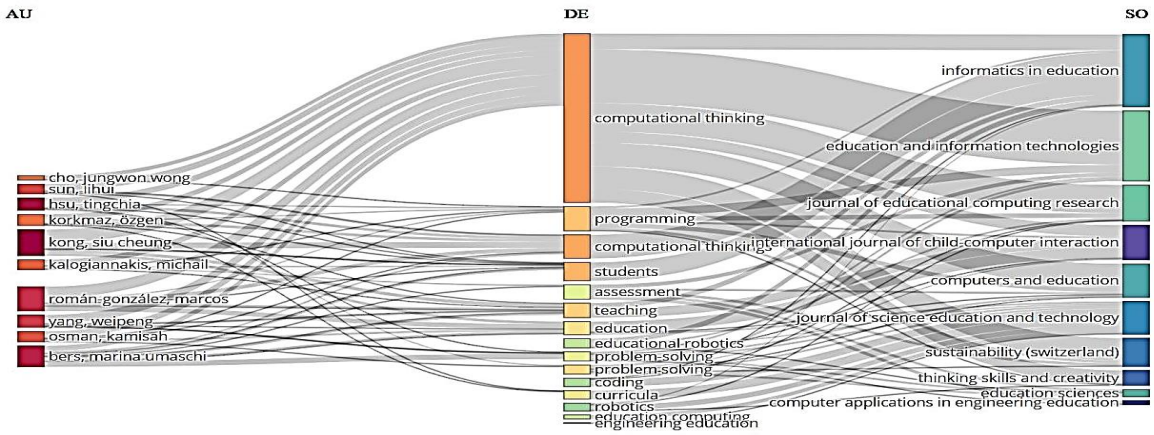
تحلیل روند زمانی تولیدات علمی، از ظهور و رشد سریع یک حوزهٔ پژوهشی نوپا حکایت دارد. همانطور که در شکل ۲ نمایان است، این حوزه پس از سال ۲۰۱۵ شاهد یک رشد نمایی در تعداد انتشارات بوده که در سال ۲۰۲۳ به اوج خود می‌رسد. این توسعه شتابان، که نشانگر اهمیت فزایندهٔ موضوع است، ضرورت یک ارزیابی نقّادانه برای یکپارچه‌سازی و سنجش کیفیت آثار تولید شده در این بازهٔ زمانی کوتاه را آشکار می‌سازد.



شکل ۳. توزیع جغرافیایی تولیدات علمی در حوزه سنجش تفکر محاسباتی (منبع: یافته‌های تحقیق)

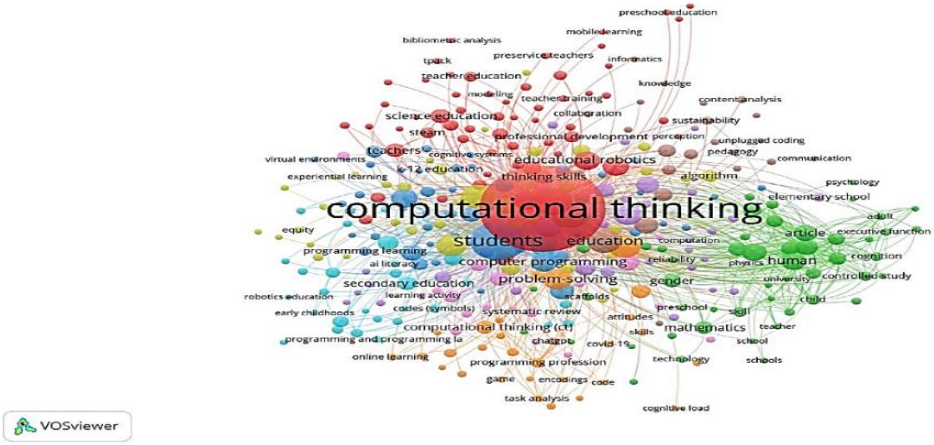
تحلیل توزیع جغرافیایی تولیدات علمی، تمرکز بالای پژوهش در مناطق خاصی از جهان را نمایان می‌سازد. نقشهٔ تولید علم کشورها (شکل ۳) نشان می‌دهد که عمدهٔ پژوهش‌ها در ایالات متحده آمریکا و چین متمرکز شده‌اند و پس از آن، کشورهایی در اروپا (مانند اسپانیا و ترکیه) و شرق آسیا (مانند تایوان) در رتبه‌های بعدی قرار دارند. این تمرکز، اگرچه بیانگر وجود زیرساخت‌های پژوهشی قوی است، اما چالش مهمی را در زمینه روایی محتوایی[1] ابزارهای سنجش ایجاد می‌کند. ابزارهای توسعه‌یافته در این بسترها ممکن است قابلیت تعمیم و کاربرد مستقیم در زمینه‌های فرهنگی و آموزشی متفاوتی مانند ایران را نداشته باشند.

1 Contextual Validity



شکل ٤. ارتباط میان نویسندگان، کلیدواژه‌ها و نشریه‌ها در حوزهٔ سنجش تفکر محاسباتی (منبع: یافته‌های تحقیق)

تحلیل نمودار سه‌وجهی (شکل ۴)، ارتباط بین نویسندگان، کلیدواژه‌ها و مجلات را ترسیم می‌کند، تصویری یکپارچه از ساختار فکری و اجتماعی این حوزهٔ پژوهشی ارائه می‌دهد. این نمودار نشان می‌دهد که نویسندگان پرکاری مانند Cho، Roman-González، Hsu و Sun آثار خود را که عمدتاً حول مفاهیم فناورانه و کاربردی متمرکز است، در مجلات اصلی حوزه فناوری آموزشی مانند Computers & Education، Journal of Educational Computing Research و Education and Information Technologies منتشر می‌کنند. این همسویی قدرتمند میان جامعهٔ پژوهشگران و مجلات فناورانه، در کنار غیبت معنادار مجلات تخصصی روان‌سنجی، الگویی را آشکار می‌سازد که در آن، فرایند توسعه و ارزیابی ابزارها در چهارچوب فناوری آموزشی محصور مانده و استانداردهای دقیق و تخصصی حوزه روان‌سنجی به شکل نظام‌مند در آن ادغام نشده است.



شکل ۵. ساختار مفهومی حوزهٔ تفکر محاسباتی بر اساس تحلیل هم‌رخدادی کلیدواژه‌ها (منبع: یافته‌های تحقیق)

تحلیل ساختار مفهومی حوزهٔ تفکر محاسباتی از طریق نقشه هم‌رخدادی کلیدواژه‌ها (شکل ۵)، شکافی عمیق میان توسعهٔ ابزار و ویژگی‌های روان‌سنجی را آشکار می‌سازد. شبکه حول یک هستهٔ قدرتمند متشکل از گره‌های (دانش‌آموزان و آموزش) سازمان یافته است که خوشه‌های بزرگ و متراکم کاربردهای آموزشی (قرمز) و ابزارهای برنامه‌نویسی (آبی) بر آن مسلط هستند. گره‌های عظیمی چون رباتیک آموزشی و آموزش پایه (تا پایان دبیرستان)، نشان‌دهندهٔ تمرکز غالب پژوهش‌ها بر کاربست آموزشی مفاهیم هستند. در مقابل، مفاهیم بنیادین سنجش به حاشیه رانده شده‌اند: گره پایایی بسیار کوچک و در موقعیتی پیرامونی قرار دارد و واژه کلیدی روایی، اساساً در نقشه غایب است. این عدم توازن بصری و شکاف ساختاری، شاهدی قدرتمند بر این مدعاست که جامعهٔ علمی، سرمایه‌گذاری بر پرسش‌های «چگونه و با چه ابزاری آموزش دهیم» را بر پرسش‌های بنیادین «چگونه معتبر بسنجیم و با چه دقتی» مقدم شمرده است.

۲. تحلیل نقّادانه ابزارهای سنجش تفکر محاسباتی

تحلیل علم‌سنجی صرفاً یک تصویر کلان ارائه نمی‌دهد، بلکه یک لنز تحلیلی برای ارزیابی نقّادانه ابزارهای موجود فراهم می‌کند. رشد سریع حوزه، تمرکز جغرافیایی پژوهش‌ها، و به‌ویژه عدم توازن مفهومی میان کاربست آموزشی و دقت روان‌سنجی، این پرسش کلیدی را مطرح می‌کند: آیا این شکاف‌ها و سوگیری‌های ساختاری در سطح خُرد و در طراحی و اعتبارسنجی خود ابزارها نیز قابل مشاهده‌اند؟ این بخش با تحلیل عمیق ۸ ابزار کلیدی (که مشخصات آن‌ها در جدول ۱ آمده است) به این پرسش پاسخ می‌دهد.

جدول ۱. ویژگی‌های روان‌سنجی و ساختاری ابزارهای منتخب سنجش تفکر محاسباتی در دانش‌آموزان ابتدایی

نام ابزار (اختصار)	نویسنده	سال	گروه هدف	ساختار و اجرا	مؤلفه‌های مورد سنجش	ویژگی‌های روان‌سنجی
CT-test	Roman-González et al	۲۰۱۷	پایهٔ پنجم تا دهم (۱۱–۱۶ سال)	۲۸ سوال (۴ گزینه‌ای) آنلاین/کاغذی ۴۵دقیقه	جهت‌ها، حلقه‌ها، شرطی ها، توابع ساده	روایی محتوایی تأیید شده روایی سازه روایی ملاکی (r=) پایایی درونی مناسب (۰/ ۷۹α≈)
TechCheck	Relkin et al	۲۰۲۰	پایهٔ اول و دوم (۵–۹ سال)	۱۵ سوال (۴ گزینه + ۲ پرسش تمرینی) آنلاین ۱۳دقیقه	الگوریتم‌ها، ماژولارسازی، ساختارهای کنترلی، بازنمایی، سخت‌افزار/نرم‌افزار، اشکال‌زدایی	روایی محتوایی تأیید شده روایی ملاکی متوسط(r=0.53) پایایی درونی=α ۰/۶۸
BCTt	Zapata-Cáceres et al	۲۰۲۰	پایهٔ اول تا چهارم (۵ تا ۱۰ سال)	۲۵ سؤال (۳و۴ گزینه‌ای)، مدادکاغذی ۴۰دقیقه	توالی، حلقه ساده، حلقه تودرتو، شرطی	روایی محتوایی تأیید شده ۰/۸۲۴α = بازآزمون rs=0.93 (بسیار پایا)
ACES	Parker et al	۲۰۲۱	پایهٔ سوم تا پنجم	۱۰ سوال بسته پاسخ، آنلاین ۱۷دقیقه	توالی‌ها، دستورهای تکرار	روایی محتوا قوی؛ شواهد سازه و ملاکی محدود α=0.686
CTA-CES	Li & Liu	۲۰۲۱	پایهٔ سوم تا ششم	۲۵ سوال (۴ گزینه‌ای) آنلاین ۴۵دقیقه	انتزاع، تفکر الگوریتمی، تجزیه، ارزیابی، الگو و تعمیم	روایی محتوایی تأیید شده پایایی همسانی درونی=α0.76
cCTt	El-Hamamsy et al	۲۰۲۲	پایهٔ سوم تا ششم (۸–۱۲ سال)	۲۵ سوال چندگزینه‌ای تصویری مدادکاغذی ۳۵دقیقه	توالی، حلقه‌های ساده، حلقه‌های پیچیده، شرطی ها، حلقه شرطی	روایی و پایایی تأییدشده α ≈ 0.82–0.85
CTtLP	Zhang & Wong	۲۰۲۳	پایهٔ اول تا سوم (۶–۱۰ سال)	۲۷ سوال ، مدادکاغذی ۶۰ دقیقه	توالی دستورات، جهت‌ها، حلقه‌ها ، شرطی‌ها	روایی محتوایی تأییدشده روایی ملاکی متوسط (r=0.44) پایایی درونی بالا (α=0.87) ثبات بازآزمون مناسب (ICC=0.76) تحلیل
CTTC	Zapata-Diaz et al	۲۰۲۴	پایهٔ اول تا سوم (۶–۹ سال)	۴۰ سوال چندگزینه‌ای تصویری مدادکاغذی ۴۰دقیقه	مهارت‌های فضایی (جهت یابی، مکان، دوران فضایی) + مهارت‌های محاسباتی (توالی دستورات، چرخه‌ها)	روایی محتوایی بالا (V-Aiken ≥ 0.70) پایایی درونی بالا (KR-20 = 0.838)

آزمون تفکر محاسباتی مبتدیان[1] (Zapata-Cáceres et al., 2020) با ۲۵ پرسش چندگزینه‌ای تصویری (ماز و بوم)، مفاهیم پایه‌ای مانند توالی، حلقه و شرط را می‌سنجد. این آزمون با طراحی بصری جذاب و کاهش وابستگی به مهارت خواندن، برای کودکان ۵ تا ۱۰ سال مناسب است. روایی محتوایی آن توسط متخصصان تأیید شده و از پایایی همسانی درونی (α =۰/۸۲۴) و بازآزمایی (rs=۰/۹۳) بسیار خوبی برخوردار است. با این حال، به دلیل سادگی، برای دانش‌آموزان بزرگتر (۹ تا ۱۲ سال) اثر سقف ایجاد کرده و پایایی آن کاهش می‌یابد. بنابراین، آزمون برای پایه‌های اول تا چهارم ابتدایی ابزار مناسبی است اما برای مقاطع بالاتر نیازمند سؤالات دشوارتر است.

آزمون سنجش محاسبات برای دانش‌آموزان ابتدایی[2] (Parker et al., 2021)، ابزاری برای دانش‌آموزان پایه‌های سوم تا پنجم ابتدایی است که تنها دو مؤلفهٔ تفکر محاسباتی یعنی توالی و تکرار را می‌سنجد. این آزمون ده پرسش بسته‌پاسخ با فرمت‌های متنوع دارد؛ برخی با قطعه کُدهای بلوکی شبیه اسکرچ و برخی به سبک رقابت ببراس برای سنجش مفاهیم غیرکُدنویسی طراحی شده‌اند. ابزار بر اساس چهارچوب مسیرهای یادگیری محاسبات روزمره شکل گرفته و روایی محتوایی آن از طریق مصاحبه‌های شناختی تأیید شده است. با این حال، در یک مطالعه مقدماتی، پایایی همسانی درونی آن ۰/۶۸ گزارش شد و شواهدی دربارهٔ روایی سازه و روایی ملاکی ارائه نشده است. بنابراین، هرچند آزمون به دلیل تنوع در نوع پرسش‌ها و سهولت اجرا (میانگین زمان ۱۷ دقیقه) اقدامی نوآورانه است، اما به دلیل پوشش محدود، پایایی پایین و فقدان شواهد کلیدی روان‌سنجی، همچنان یک مقیاس مقدماتی و آزمایشی محسوب می‌شود.

آزمون سنجش تفکر محاسباتی برای دانش‌آموزان ابتدایی چینی[3] (Li & Liu, 2021) ویژهٔ دانش‌آموزان سوم تا ششم ابتدایی در چین طراحی شد. آزمون شامل ۲۵ سؤال چندگزینه‌ای تصویری است که مسائل را در بستر واقعی ارائه می‌دهد. این ابزار پنج مؤلفهٔ اصلی تفکر محاسباتی (انتزاع، الگوریتم، تجزیه، ارزیابی و الگو/تعمیم) را پوشش می‌دهد. از نظر روایی محتوایی توسط متخصصان و مصاحبه‌های شناختی تأیید شد. روایی سازه از طریق همبستگی منفی و معنادار با تجربهٔ برنامه‌نویسی و مهارت‌های دیجیتال پشتیبانی شد و روایی ملاکی آن با همبستگی مثبت با توانایی استدلال (r=۰/۴۷)، فضایی (r=۰/۴۳) و خواندن (r=۰/۵۴) نشان داده شد. پایایی آزمون نیز با آلفای کرونباخ (α=۰/۷۶) در سطح قابل قبول گزارش شد. با وجود پوشش جامع و شواهد روان‌سنجی متنوع، محدودیت اصلی این ابزار در تعمیم‌پذیری فرهنگی و پایایی متوسط آن است؛ بنابراین، اعتبار آن برای استفاده در بسترهای غیرچینی نیازمند بررسی‌های بیشتر است.

آزمون شایستگی تفکر محاسباتی[4] (El-Hamamsy et al., 2022) برای دانش‌آموزان پایهٔ سوم تا ششم طراحی شده و مفاهیمی چون توالی، حلقه و شرط را می‌سنجد. روایی محتوایی و سازهٔ آن با نظر متخصصان و تحلیل عاملی تأیید شده و پایایی همسانی درونی بسیارخوبی (بین ۰/۸۲ تا ۰/۸۵) دارد که نشان‌دهندهٔ ثبات مطلوب ابزار است. نقطهٔ قوت برجستهٔ این ابزار، استفاده از تحلیل‌های روان‌سنجی پیشرفته است که دشواری مناسب و نبود سوگیری جنسیتی را نشان می‌دهد. با این حال، محدودیت اصلی آن، سادگی بیش از حد برخی سؤال‌ها برای پایه‌های بالاتر است که منجر به «اثر سقف» می‌شود. این مسئله کارایی آزمون را برای سنجش دانش‌آموزان توانمندتر کاهش می‌دهد و نشان‌دهندهٔ نیاز به افزودن پرسش‌های دشوارتر است.

آزمون تفکر محاسباتی برای دورهٔ اول ابتدایی[5] (Zhang & Wang, 2023) ، با طراحی کودک‌پسند و ۲۷ پرسش چندگزینه‌ای، ابزاری برای سنجش مؤلفه‌های بنیادی مانند توالی، جهت‌ها، حلقه‌ها و شرطی‌ها در پایه‌های اول تا سوم است. از منظر روان‌سنجی، این ابزار شواهد قوی ارائه می‌دهد: روایی محتوایی آن تأیید شده، پایایی همسانی درونی بالا (α=۰/۸۷) و ثبات بازآزمون مطلوب (ICC=۰/۷۶) همچنین، روایی ملاکی متوسط (r=۰/۴۴) با عملکرد در درس اسکرچ و تحلیل IRT نشان می‌دهد که پرسش‌ها دارای دشواری متنوع و قدرت تمایز بالا هستند. با این حال، ماهیت کاملاً چندگزینه‌ای آن، توانایی سنجش فرایندهای حل مسئله را کاهش می‌دهد و تمرکز انحصاری بر مفاهیم، باعث نادیده گرفتن شیوه‌هایی مانند اشکال‌زدایی و نگرش‌هایی چون پشتکار می‌شود. این محدودیت‌ها نشان می‌دهد که آزمون گرچه در سنجش دانش مفهومی موفق است، اما در ارزیابی توانایی عملی و چندبُعدی دانش‌آموزان با کاستی‌هایی مواجه است.

---

1 Beginners' Computational Thinking Test
2 Assessing Computation for Elementary Students
3 Computational Thinking Test for Upper Primary School Students
4 Computational Thinking Competency Test
5 Computational Thinking Test for Lower Primary Students

آزمون تفکر محاسباتی برای کودکان[1] (Zapata-Caceres & Martín-Barroso, 2024) برای پایه‌های اول تا سوم ابتدایی طراحی شده و شامل ۴۰ سؤال چندگزینه‌ای تصویری است. این ابزار پنج بُعد اصلی شامل مهارت‌های فضایی (جهت‌یابی، دوران) و مهارت‌های محاسباتی (توالی، چرخه) را در کنار مفاهیمی چون تفکر منطقی و شناسایی الگو می‌سنجد. روایی محتوایی آن توسط متخصصان و محاسبهٔ شاخص آیکن (Aiken's V>۰/۷۰) تأیید شده و پایایی همسانی درونی بالایی (KR-20=۰/۸۳۸) دارد. طراحی جذاب تصویری و دشواری متناسب، آن را به ابزاری مناسب برای کودکان تبدیل کرده است. با این حال، تمرکز بر قالب چندگزینه‌ای یک محدودیت ذاتی است، آزمون در سنجش مهارت‌های عملی و نگرشی مانند اشکال‌زدایی و پشتکار، با کاستی مواجه است. در نتیجه، آزمون برای ارزیابی مفاهیم بنیادی قدرتمند است، اما برای یک تصویر کامل، باید در کنار ابزارهای سنجش عملکردی[2] به کار رود.

در مجموع، تحلیل ابزارهای مرورشده نشان می‌دهد که هیچ ابزار واحدی کامل نیست و هر یک با مجموعه‌ای از نقاط قوت و ضعف همراه است. این یافته‌ها در بخش بعد به صورت موضوعی مورد بحث و بررسی قرار خواهند گرفت.

بحث

یافته‌های این پژوهش، تصویری چندلایه و پیچیده از وضعیت سنجش تفکر محاسباتی ارائه می‌دهد. برای درک دقیق‌تر این یافته‌ها، بحث در سه محور اصلی سازماندهی شده است:

۱. شکاف ساختاری: همسویی معنادار میان پارادایم غالب و کیفیت روان‌سنجی

تحلیل همزمان نقشه‌های علم‌سنجی و ویژگی‌های ابزارها، الگویی قابل تأمل را آشکار می‌سازد. در حالی که خوشه‌های «کاربردهای آموزشی» و «ابزارهای برنامه‌نویسی» بر نقشه علمی مسلط هستند، گره‌های مربوط به «روایی» و «پایایی» در حاشیه قرار گرفته‌اند. این یافته‌ی کلان، نمود عینی خود را در تحلیل مقاله‌های منتخب نشان می‌دهد؛ جایی که ضعف‌های روش‌شناختی در ابزارهای توسعه‌یافته فراگیر است.

به‌طور مشخص، اغلب مطالعات به گزارش ضرایب سنتی چون آلفای کرونباخ بسنده کرده‌اند که بر پیش‌فرض‌های سخت‌گیرانه‌ای استوار است که به ندرت در آزمون‌های شناختی برقرار است. این در حالی است که متخصصان روان‌سنجی امروزه استفاده از شاخص‌های دقیق‌تری مانند «امگای مک‌دونالد» را توصیه می‌کنند (Hayes & Coutts, 2020؛ McNeish, 2018). به همین ترتیب، غفلت فراگیر از مفهوم انصاف و عدم بررسی کارکرد افتراقی سؤالات (DIF) به استثنای یک مورد (El-Hamamsy et al., 2022) نشان می‌دهد که این حوزه از استانداردهای مدرن آزمون‌سازی فاصله دارد. اگرچه نمی‌توان با قطعیت ادعای رابطهٔ علّی کرد، اما همسویی این دو پدیده نشان می‌دهد که در اکوسیستم پژوهشی فعلی، اولویت‌دهی به توسعهٔ سریع ابزار، احتمالاً پژوهشگران را از تمرکز بر این پیچیدگی‌های دقیق روان‌سنجی بازداشته است.

۲. عدم توازن مفهومی: سیطره مفاهیم بر فرایندها

یافته‌های مرور حاضر نشان داد که اکثر ابزارها بر مفاهیم محاسباتی (مانند توالی و حلقه) تمرکز دارند و ابعاد فرایندی (مانند انتزاع و تعمیم) و نگرشی را کمتر پوشش می‌دهند. این محدودیت مفهومی را می‌توان بازتابی از «تسهیل‌گری در سنجش» دانست؛ چراکه سنجش دانشِ الگوریتمی در قالب آزمون‌های چندگزینه‌ای (مانند CTt) بسیار ساده‌تر از سنجش فرایندهای پیچیدهٔ حل مسئله است. این رویکرد تقلیل‌گرایانه، اگرچه سنجش را در مقیاس وسیع ممکن می‌سازد، اما خطر نادیده گرفتن ماهیت چندبُعدی تفکر محاسباتی را به همراه دارد؛ موضوعی که شوت و دیگران (Shute et al., 2017) نیز بر آن تأکید داشته‌اند.

۳. چالش تعمیم‌پذیری فرهنگی و زبانی

تمرکز جغرافیایی پژوهش‌ها در کشورهایی مانند آمریکا و چین، که در نقشه تولید علم مشخص شد، چالش جدی «روایی فرهنگی» را مطرح می کند. ابزارهای توسعه‌یافته در این بافت‌های جغرافیایی، غالباً بر مبنای پیش‌فرض‌های آموزشی و زیرساختی خاصی (مانند دسترسی فراگیر به فناوری های دیجیتال یا ادغام برنامه‌نویسی در برنامه درسی رسمی) طراحی شده‌اند که ممکن است در سایر نظام‌های آموزشی، به‌ویژه ایران، صادق نباشد. فقدان ابزارهای بومی استاندارد در میان مقالات بررسی‌شده، نشان می‌دهد که صرفِ ترجمه ابزارهای موجود کافی نیست و نیاز مبرمی به بومی سازی و انطباق فرهنگی مادّه‌های آزمون‌ها جهت تضمین روایی آزمون در پژوهش‌های ملّی احساس می‌شود.

---

1 Computational Thinking test for Primary Education
2 performance-based

نتیجه‌گیری

مرور نقّادانه حاضر نشان داد که چالش‌های موجود در سنجش تفکر محاسباتی، فراتر از ضعف‌های تکنیکیِ موردی بوده و ریشه در عدم توازن ساختاریِ کل حوزهٔ پژوهشی دارد. جامعهٔ علمی این حوزه، که عمدتاً در فضای غالب فناوری آموزشی فعالیت می‌کند، ابزارهایی را توسعه داده که اگرچه از نظر آموزشی کاربردی‌اند، اما در سنجش معتبر، عادلانه و جامع یک سازهٔ پیچیده با کاستی‌های جدی مواجه‌اند. سهم اصلی این پژوهش، ارائه یک تحلیل تشخیصی است که با پیوند زدن تحلیل کلان (نقشه علمی) به تحلیل خرد (ویژگی‌های روان‌سنجی)، ریشه‌های سیستمی این مشکلات را آشکار می‌سازد.

بر مبنای تحلیل‌های انجام شده، این پژوهش یک نقشه راه سه مرحله‌ای را برای گذار از وضعیت موجود و بلوغ حوزهٔ سنجش ترسیم می‌کند. در گام نخست، ارتقای بنیان‌های روان‌سنجی از طریق عبور از شاخص‌های سنتی و به‌کارگیری الزامی روش‌های مدرن مانند مدل‌های نظریهٔ سؤال–پاسخ (IRT)، شاخص پایایی امگا و تحلیل‌های انصاف (DIF) یک اولویت فوری محسوب می‌شود. همگام با این تحول روش‌شناختی، گام دوم بر توسعهٔ مدل‌های ترکیبی و کل‌نگر تمرکز دارد؛ رویکردی که در آن آزمون‌های استاندارد با روش‌های مکمل مانند ارزیابی مبتنی بر پروژه و پوشه‌کار تلفیق می‌شوند تا تصویری جامع‌تر از شایستگی دانش‌آموز ارائه گردد؛ ضرورتی که (Tang et al., 2020) نیز بر آن صحه گذاشته‌اند. در نهایت، گام سوم بر بومی‌سازی هوشمند ابزارها معطوف است؛ بدین معنا که توسعه و هنجاریابی ابزارها باید متناسب با شکاف‌های فرهنگی و مقتضیات برنامه درسی ملّی (به‌ویژه برای دورهٔ دوم ابتدایی) صورت گیرد تا از روایی و کارایی آن‌ها در بافت بومی اطمینان حاصل شود.

تحقق این نقشه راه، پیامدهای کاربردی مهمی نیز برای نظام آموزشی خواهد داشت. گذار به سمت ابزارهای استاندارد و بومی، به معلمان و سیاست‌گذاران اجازه می‌دهد تا به جای تکیه بر شواهد پراکنده، اثربخشی برنامه‌های درسی تفکر محاسباتی را با دقت رصد کرده و مداخلات آموزشی را بر مبنای داده‌های معتبر طراحی کنند. در نهایت، این پژوهش پیشنهاد می‌کند که آینده مطالعات ابزارسازی باید از تولید انبوه ابزارهای مشابه به سمت ارتقای کیفیت و استانداردسازی ابزارهای موجود تغییر جهت دهد.

این پژوهش با وجود رویکرد نظام‌مند، با محدودیت‌هایی مواجه است که باید در تعمیم نتایج درمد نظر قرار گیرند. نخست، دامنهٔ جستجو به مقالات منتشرشده به زبان انگلیسی در پایگاه‌های منتخب محدود بود و ادبیات پژوهشی سایر زبان‌ها و همچنین بخشی از ادبیات خاکستری (نظیرگزارش‌های فنی و پایان‌نامه‌ها) را پوشش نداد. همچنین، تمرکز اصلی این مرور بر ابزارهای استاندارد کمّی (آزمون‌ها و پرسشنامه‌ها) معطوف بود و سایر روش‌های ارزیابی کیفی مانند مصاحبه‌های بالینی یا تحلیل تکالیف باز–پاسخ به صورت عمیق بررسی نشدند. در نهایت، عمق تحلیل نقّادانه تا حدودی تحت تأثیر کیفیت ناهمگون گزارش‌دهی در مطالعات اولیه قرار داشت؛ به طوری که فقدان جزئیات کامل روان‌سنجی در برخی منابع، امکان مقایسهٔ تطبیقی دقیق‌تر را در مواردی با چالش مواجه می‌ساخت.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

این پژوهش کد اخلاق را به شماره IR.ALZAHRA.REC.1404.081 از کمیتهٔ اخلاق دانشگاه الزهرا دریافت کرده است.

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده‌اند و این موضوع مورد تأیید همه آنهاست.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

سپاسگزاری

این مقاله مستخرج از رساله دکتری نویسندهٔ اول در دانشگاه الزهرا است. بدین‌وسیله از معاونت محترم پژوهشی دانشگاه و همچنین از داوران محترم نشریه جهت ارائه نظرهای ساختاری و علمی ارزشمند که موجب غنای این اثر شد، تشکر و قدردانی می‌شود.

منابع

رزبان، س.، شاه حسینی، س.، و باقری، م. (۱۳۹۹). تأثیر آموزش برنامه نویسی با نرم افزار اسکرچ بر مهارت حل مسئله دانش آموزان. *آموزش و ارزشیابی (علوم تربیتی)*، ۱۳(۴۹)، ۹۳-۱۰۷. https://doi.org/10.30495/jinev.2020.673920

رمضانی، ع.، و شریفی، م. (۱۴۰۳). مدل‌سازی تأثیر شکاف دیجیتال، تفکر رایانشی (محاسباتی) و جذب شناختی بر سواد هوش مصنوعی در بین دانشجو معلمان. *مطالعات آموزش و یادگیری*، ۱۶(۲)، ۲۵۹-۲۷۸. http://noo.rs/xp3h9

روحی، م.، دهقانی، م.، و خطاط، م. (۱۴۰۲). شناسایی مؤلفه‌های تفکر محاسباتی در مهارت‌های یادگیری دوره ابتدایی. *مطالعات آموزش و یادگیری*، ۱۵(۱)، ۱۶ – ۴۵. https://doi.org/10.22099/jsli.2023.7180

شایان، م.، و یافتیان، ن. (۱۳۹۷). ارزیابی عملکرد دانش‌آموزان پایه دهم و معلمان متوسطه اول در آزمون سواد ریاضی. *پژوهش در نظام‌های آموزشی*، ۱۲(۴۳)، ۶۵-۹۴. https://doi.org/10.22034/jiera.2018.83764

ضیاءنژاد شیرازی، آ.، کوروش نیا، م.، سهرابی، ن.، و بقولی، ح. (۱۴۰۱). تدوین معادله افتراقی عوامل فردی موثر بر عملکرد در آزمون تیمز ۲۰۱۹ پایه چهارم در دانش آموزان با عملکرد بالا و پایین. *پژوهش در نظام‌های آموزشی*، ۱۶(۵۸)، ۱۴۱-۱۲۹. https://doi.org/10.22034/jiera.2022.167791

مومنی کوهستانی، ا.، و نژادحسینیان، ی. (۱۴۰۱، آذر). تاثیر به کار گیری نرم افزار اسکرچ بر خلاقیت دانش آموزان. *هفتمین کنفرانس ملی رویکردهای نوین در آموزش و پژوهش*، محمودآباد. https://civilica.com/doc/1619732

References

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 Annual Meeting of the American Educational Research Association*, Vancouver, Canada, 1, 25. https://web.media.mit.edu/~kbrennan/files/Brennan_Resnick_AERA2012_CT.pdf

El-Hamamsy, L., Zapata-Cáceres, M., Barroso, E. M., Mondada, F., Zufferey, J. D., & Bruno, B. (2022). The competent computational thinking test: Development and validation of an unplugged computational thinking test for upper primary school. *Journal of Educational Computing Research*, *60*(7), 1818-1866. https://doi.org/10.1177/07356331221081753

Fagerlund, J., Häkkinen, P., Vesisenaho, M., & Viiri, J. (2021). Assessing computational thinking in a programming course for pre-service teachers. *Informatics in Education*, 20(2), 213–236. https://doi.org/10.15388/infedu.2021.10

González, M. J. C., Pérez, J. M., & Chas-Apiñeiro, M. (2017). A tool for the assessment of computational thinking skills in primary school students. *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'17)*, 1–7. Association for Computing Machinery. https://doi.org/10.1145/3144826.3145404

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher*, 42(1), 38–43. https://doi.org/10.3102/0013189X12463051

Haseski, H. İ., Ilić, U., & Tuğtekin, U. (2018). The relationship between computational thinking skills, and programming self-efficacy and problem-solving skills in the programming process. *Journal of Educational Computing Research*, 56(6), 848–871. https://doi.org/10.1177/0735633117718298

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Communication Methods and Measures*, 14(1), 1–24.https://doi.org/10.1080/19312458.2020.1718629

Hsu, T. C., & Chen, M. S. (2025). Effects of students using different learning approaches for learning computational thinking and AI applications. *Education and Information Technologies*, 30(6), 7549–7571. https://doi.org/10.1007/s10639-024-12933-w

Li, Y., & Liu, K. (2021). Development and validation of a computational thinking test for upper primary school students. *Journal of Educational Computing Research*, 59(3), 457–483. https://doi.org/10.1177/0735633120967347

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. https://doi.org/10.1037/met0000147

Momeni Kohestani, A. M., & Nezhadhosseinian, Y. (2022, December). The effect of using Scratch software on students' creativity. *7th National Conference on New Approaches in Education and Research*, Mahmudabad. (in Persian) https://civilica.com/doc/1619732

Ocampo, J. L. D., Herman, J., & Suson, R. (2024). A systematic review of computational thinking assessment tools in K-12 education. *Education and Information Technologies*, 29(1), 329–358. https://doi.org/10.1007/s10639-023-11756-5

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Parker, M. C., Chen, Y., O'Neal, L. J., Wagh, A., Tucker-Cuthbertson, A. J., & Century, J. (2021). Assessing computation for elementary students (ACES): An assessment of sequences and repeating patterns. *Proceedings of the 2021 ACM Conference on Innovation and Technology in Computer Science Education V. 1*, 518–524. Association for Computing Machinery. https://doi.org/10.1145/3430665.3456382

Ramezani, A., & Sharifi, M. (2024). Modeling the effect of digital divide, computational thinking and cognitive absorption on artificial intelligence literacy among student teachers. *Journal of Studies in Learning and Instruction*, 16(2), 259–278. (in Persian) http://noo.rs/xp3h9

Razban, S., Shahhosseini, S., & Bagheri, M. (2020). The effect of teaching programming with Scratch software on students problem solving skills. *Journal of Instruction and Evaluation*, 13(49), 93–107. (in Persian) https://doi.org/10.30495/jinev.2020.673920

Relkin, E., de Ruiter, L., & Bers, M. U. (2020). TechCheck: Development and validation of an unplugged computational thinking assessment for early primary education. *Journal of Science Education and Technology*, 29, 482–498. https://doi.org/10.1007/s10956-020-09831-x

Rouhi, M., Dehghani, M., & Khatat, M. (2023). Identifying components of computational thinking in elementary school learning skills. *Journal of Studies in Learning and Instruction*, 15(1), 16–45. (in Persian) https://doi.org/10.22099/jsli.2023.7180

Selby, C., & Woollard, J. (2013). *Computational thinking: The developing definition*. University of Southampton. http://eprints.soton.ac.uk/356481/

Shayan, M., & Yaftian, N. (2018). Evaluation of the performance of 10th grade students and junior high school teachers in mathematical literacy test. *Journal of Research in Educational Systems*, 12(43), 65–94. (in Persian) https://doi.org/10.22034/jiera.2018.83764

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Stewart, W. H., et al. (2021). The effects of collaborative coding on upper elementary students' computational thinking. *Computer Science Education*, 31(4), 432–465. https://doi.org/10.1080/08993408.2020.1866931

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. Computers & Education, 148, 103798. https://doi.org/10.1016/j.compedu.2019.103798

Tang, X., Wang, M., & Qian, Y. (2020). A systematic review of computational thinking assessment for K-12 students. *Journal of Science Education and Technology*, 29, 626–641. https://doi.org/10.1007/s10956-020-09840-w

Tikva, C., & Tambouris, E. (2021). Mapping computational thinking through programming in K-12 education: A systematic literature review. *Computers & Education*, 162, 104082. https://doi.org/10.1016/j.compedu.2020.104082

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. https://doi.org/10.1145/1118178.1118215

Yadav, A., et al. (2022). Computational thinking in education: A pedagogical perspective. *Handbook of Research on Teaching*. https://doi.org/10.3102/978-0-935302-48-6_8

Zapata-Cáceres, M., & Martín-Barroso, E. (2024). Validation of the computational thinking test for primary education students (CTt-PE). *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 19(1), 105–115. https://doi.org/10.1109/RITA.2024.3364402

Zapata-Cáceres, M., Martín-Barroso, E., & Román-González, M. (2020). Bebras and CTt-v2: Two instruments for the assessment of computational thinking. *Journal of Information Technology Research (JITR)*, 13(4), 1–20. https://doi.org/10.4018/JITR.2020100101

Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*, 141, 103607. https://doi.org/10.1016/j.compedu.2019.103607

Zhang, L., & Wang, Q. (2023). Development and validation of a computational thinking test for lower primary students. *Journal of Computer Assisted Learning*, 39(1), 116–129. https://doi.org/10.1111/jcal.12736

Zianezhad Shirazi, A., Kouroshnia, M., Sohrabi, N., & Baghouli, H. (2022). Formulating differential equation of individual factors affecting performance in TIMSS 2019 fourth grade in high and low performance students. *Journal of Research in Educational Systems*, 16(58), 129–141. (in Persian) https://doi.org/10.22034/jiera.2022.167791