MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2026, 22(2), em2775
ISSN:1305-8223 (online)
OPEN ACCESS Review Article https://doi.org/10.29333/ejmste/17827

Cognitive diagnostic assessment in educational contexts: Global trends, methodological innovations, and emerging direction

Fitria Herliana [1,2], A. Halim [2*], Riza Andriani [3], Burhanuddin Yasin [4], Mailizar [5]

[1] Student of Doctoral Education Department, Universitas Syiah Kuala, INDONESIA
[2] Lecturer of Physics Education Department, Universitas Syiah Kuala, INDONESIA
[3] Lecturer of Physics Education Department, Universitas Malikussaleh, INDONESIA
[4] Lecturer of English Education Department, Universitas Syiah Kuala, INDONESIA
[5] Lecturer of Mathematics Education Department, Universitas Syiah Kuala, INDONESIA

Received 16 October 2025 ▪ Accepted 22 January 2026

Abstract

Cognitive diagnostic assessment (CDA) provides precise insights into students' cognitive strengths and weaknesses. This systematic literature review analyzed 56 Scopus-indexed articles (2015-2025) using the theory-context-method framework. CDA is predominantly applied in mathematics, science, and language education, with significant contributions from China, USA, and Malaysia. While classical DINA and G-DINA models remain prevalent, recent studies integrate machine learning for Q-matrix validation. Quantitative approaches dominate (76.8%), revealing a gap between technical sophistication and practical implementation. The review emphasizes CDA's potential for personalized learning and evidence-based policy, recommending future research to adopt longitudinal designs, expand interdisciplinary integration, and bridge diagnostic insights with pedagogical practice.

Keywords: CDA, cognitive diagnostic model, Q-matrix, personalized learning, diagnostic psychometrics

INTRODUCTION

Cognitive diagnostic assessment (CDA) has increasingly attracted scholarly attention as an innovative approach that enables educators to diagnose students' cognitive strengths and weaknesses with greater precision. By employing cognitive diagnostic models (CDMs), teachers can identify specific skills or attributes that learners have or have not mastered, thus providing a foundation for more targeted instruction (Tatsuoka, 2009; Templin & Henson, 2010). Since the early 2010s, research in this field has expanded rapidly, ranging from early attempts that utilized artificial neural networks for diagnostic classification (Cui et al., 2016) to applications across mathematics (Chin et al., 2021b; Wu, 2019), reading (Li et al., 2021; Ranjbaran & Alavi, 2017), and science education (Zhou & Traynor, 2022). International comparative assessments such as TIMSS and PISA further highlight the relevance of CDA/CDMs in understanding cross-national variations in student learning trajectories (Wu et al., 2020; Yamaguchi & Okada, 2018).

Nevertheless, despite significant advances, several challenges remain unresolved. Conventional psychometric models such as DINA and DINO, though widely applied, are often limited in handling complex learning data (Ravand & Robitzsch, 2018; Templin & Henson, 2006). The validation of Q-matrix—a critical component of CDA/CDMs—has been identified as a persistent issue that directly affects diagnostic accuracy (Qin & Guo, 2023). Moreover, existing reviews of CDA/CDM research tend to focus on isolated aspects, either model development, Q-matrix validation, or specific subject domains, without offering an integrative synthesis across contexts and methodologies (Lin et al., 2020; Nájera et al., 2019). This fragmentation underscores the lack of a comprehensive understanding of how CDA/CDMs evolve when combined with emerging technologies and applied in diverse educational settings.

✉ fitriaherliana@usk.ac.id ✉ abdul.halim@usk.ac.id **(*Correspondence)** ✉ rizaandriani@unimal.ac.id
✉ burhanyasin@usk.ac.id ✉ mailizar@usk.ac.id

Contribution to the literature

- This study introduces the theory-context-method (TCM) framework as an integrative analytical lens for systematically mapping CDA research, enabling comprehensive synthesis across theoretical, contextual, and methodological dimensions. The review provides the first comprehensive temporal-geographical mapping of CDA research (2015-2025), revealing critical knowledge transfer patterns between Western theoretical contributions and Asian empirical applications, alongside systematic quantification of the theory-practice gap.
- Methodologically, this study establishes reproducible standards for future CDA reviews through rigorous preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocols with documented inter-rater reliability (κ = 0.85-0.88) and explicit quality assessment procedures.
- The findings reframe CDA from a specialized psychometric tool into a transformative framework for personalized learning, offering evidence-based directions for interdisciplinary integration, longitudinal designs, and practitioner-engaged research that bridges diagnostic sophistication with pedagogical action.

At the same time, the body of literature shows several promising trends. Recent studies have integrated machine learning and neural networks to enhance Q-matrix validation and improve diagnostic performance (Qin & Guo, 2023; Tao et al., 2024). Others have explored CDA/CDMs in online environments, adaptive learning platforms, and large-scale international assessments, pointing to their potential in personalizing instruction and informing policy (Chin et al., 2021b; Jiang et al., 2022; Toprak-Yildiz, 2021; Wu et al., 2020). Yet, while these developments, they remain fragmented, and their implications for validity, reliability, and classroom practice are not fully consolidated. This highlights the urgency of a more holistic synthesis that captures methodological advances alongside contextual applications.

In response to these gaps, this study conducts a systematic literature review (SLR) of CDA/CDM research published 2015-2025, following the PRISMA protocol. The objectives are threefold:

(1) to profile global publication trends in CDA/CDMs, including geographic distribution, methodological approaches, and theoretical orientations,

(2) to analyze thematic developments such as Q-matrix validation, technological integration, and interdisciplinary applications, and

(3) to identify limitations in existing studies and propose directions for future research.

Guided by these aims, the review addresses the following research questions:

1. How have CDA and CDMs evolved in education over the past decade (2015-2025)?
2. What are the significant theoretical and methodological contributions in CDA/CDM research?
3. How can CDA/CDMs be applied across different educational contexts to enhance learning?
4. What gaps and limitations in CDA/CDM research need to be addressed in future studies?

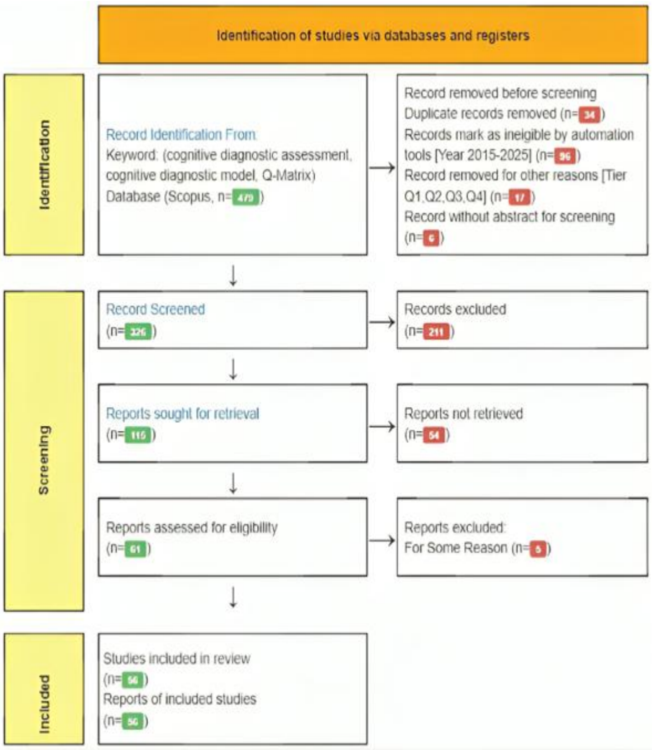

Figure 1. PRISMA flowchart (Source: Authors' own elaboration, using Watase Uake Tools)

To address these limitations, this review introduces an integrative classification framework that systematically connects methodological innovations, theoretical advancements, and contextual applications. By consolidating previously fragmented research streams, the study not only advances theoretical discourse but also provides practical implications for adaptive learning systems and evidence-based educational policy.

METHOD

Research Design

This study employed an SLR approach, following the guidelines of the PRISMA (Moher et al., 2009). PRISMA was chosen because it has been widely recognized for enhancing methodological rigor and transparency in reporting (Panic et al., 2013; Siddaway et al., 2019). The article selection process is visualized through the PRISMA flowchart in **Figure 1**.

Search Strategy and Database Selection

The literature search was conducted using the Scopus database, selected for its wide coverage, rigorous indexing standards, and comprehensive multidisciplinary scope. Search strings were designed to capture relevant CDA/CDM literature comprehensively: TITLE-ABS-KEY ("cognitive diagnostic assessment" OR "cognitive diagnostic model*" OR "CDM" OR "CDA") AND TITLE-ABS-KEY ("Q-matrix" OR "Q matrix" OR "attribute" OR "diagnostic classification"). The search was limited to publications published 2015-2025, with document type restricted to journal articles, written in English, and indexed under the subject areas of social sciences, psychology, computer science, and mathematics. The initial search was conducted on August 30, 2025, and updated on September 28, 2025, to capture the most recent publications. The final search yielded 479 records from the Scopus database.

Inclusion and Exclusion Criteria

Inclusion criteria

Studies were included if they met the following criteria:

(a) empirical studies, theoretical papers, or methodological contributions directly related to CDA or CDMs in educational contexts,

(b) published in peer-reviewed journals indexed in Scopus,

(c) focused on educational settings including K-12, higher education, or professional training,

(d) published 2015-2025, and

(e) written in English.

Exclusion criteria

Studies were excluded if they:

(a) were conference papers, book chapters, dissertations, or other non-journal publications,

(b) did not directly address CDA or models in educational contexts,

(c) were non-empirical opinion pieces without methodological contribution,

(d) were identified as duplicate publications,

(e) had inaccessible full text despite institutional access and direct author contact, or

(f) focused on non-educational domains such as clinical psychology or medical diagnosis.

Screening Process

Two independent reviewers conducted screening with structured consensus procedures. A third reviewer resolved disagreements when consensus could not be reached.

Identification and initial filtering

From 479 initial records, automated and manual filtering removed: 34 duplicate records, 96 records outside the date range (2015-2025), 17 records from non-peer-reviewed or predatory sources (identified through Scopus CiteScore verification), and 6 records without abstracts. This yielded 326 records for title and abstract screening.

Title and abstract screening

Two reviewers independently screened all 326 records based on titles and abstracts against the inclusion/exclusion criteria. Inter-rater reliability, calculated on a random sample of 100 articles, yielded $\kappa = 0.74$ (substantial agreement) (Landis & Koch, 1977). After screening, 211 records were excluded as they did not meet the inclusion criteria based on title and abstract review (primarily studies not directly addressing CDA/CDM in educational contexts, studies with tangential relevance, or studies clearly outside the review scope). This left 115 reports sought for full-text retrieval.

Full-text retrieval

Of 115 reports sought for retrieval, 54 could not be accessed despite institutional subscriptions, interlibrary loan requests, and direct author contact (28 no institutional access; 18 non-responsive authors; 8 broken links). The remaining 61 full-text reports were assessed for eligibility. Inter-rater reliability at this stage was $\kappa = 0.88$ (almost perfect agreement) (Landis & Koch, 1977).

All 61 full-text articles were assessed using a domain-specific checklist with five criteria:

(1) research design clarity,

(2) methodological rigor,

(3) data quality and transparency,

(4) results validity, and

(5) contribution to CDA field.

Each criterion was rated yes (2 points), partial (1 point), or no (0 points), with inclusion threshold ≥ 7/10 points.

The tool was piloted on 10 articles, achieving 90% agreement after refinement of operational definitions. Two reviewers independently assessed all 61 articles with inter-rater reliability $\kappa = 0.80$ (almost perfect agreement) (Landis & Koch, 1977). Of 61 assessed, 56 met the threshold (38 scored 9-10 points; 18 scored 7-8 points), and 5 were excluded for insufficient rigor or inadequate reporting.

Table 1. Journal distribution

Journal	Tier	C	TA
Frontiers in Psychology	Q1	109	14
Applied Psychological Measurement	Q1	108	6
Educational Psychology	Q1	136	6
Educational and Psychological Measurement	Q1	135	4
Studies in Educational Evaluation	Q1	50	3
Applied Sciences	Q2	35	3
Current Psychology	Q2	34	3
Journal of Computers in Education	Q1	13	2
Journal of Educational Measurement	Q1	9	2
Behavior Research Methods	Q1	3	1
British Journal of Mathematical and Statistical Psychology	Q1	55	1
Education and Information Technologies	Q1	2	1
International Journal of Listening	Q1	32	1
International Journal of Science and Mathematics Education	Q1	9	1
International Journal of Testing	Q1	3	1
Journal of Classification	Q1	11	1
Language Testing in Asia	Q1	17	1
Large-scale Assessments in Education	Q1	2	1
Physical Review Physics Education Research	Q1	0	1
PLoS ONE	Q1	27	1
The Journal of Experimental Education	Q1	23	1
Journal of Psychoeducational Assessment	Q2	1	1

Note. C: Citation & TA: Total articles

Data Extraction and TCM Framework Classification

The 56 included studies were analyzed using the TCM framework, which provides systematic classification along three dimensions: theory—theoretical models or frameworks employed; context—educational settings including subject domain, educational level, and geographic location; and method—research paradigms, analytical approaches, and data sources utilized. Two researchers independently extracted data from all 56 studies using standardized extraction forms capturing bibliographic information, theoretical framework employed, educational context, research methods, analytical approaches, and key contributions. Following data extraction, both researchers independently coded all studies according to the TCM framework using a structured coding manual with explicit operational definitions for each dimension.

The TCM coding procedures underwent pilot testing on randomly selected studies to calibrate interpretation and ensure consistency. Final inter-coder agreement across all 56 studies was 92.3% overall, with dimension-specific agreement of: theory dimension 92.9%, context dimension 96.4%, and method dimension 87.5%. The 13 disagreements were resolved through discussion achieving 100% final consensus. Complete TCM classification for all 56 studies is transparently presented in **Appendix A** enabling readers to verify coding decisions and assess classification validity.

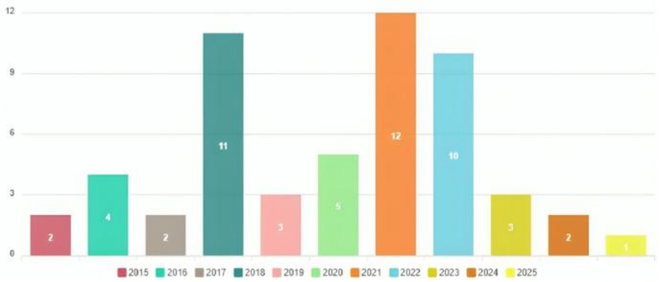

Figure 2. Year of publication (Source: Authors' own elaboration, using Watase Uake Tools)

RESULTS

This section presents findings from the systematic analysis of 56 studies on CDA and CDMs published between 2014 and 2025. Results are organized according to the TCM framework, with complete classification of all studies provided in **Appendix A**.

Overview of Included Studies

Table 1 displays the journal distribution of the 56 analyzed studies. The vast majority (49 studies) were published in Q1 journals, reflecting the academic maturity of CDA research and its international recognition. Frontiers in Psychology emerged as the leading outlet with 14 publications that bridge psychology, education, and technological innovation (Huang et al., 2022; Jia et al., 2021; Lin et al., 2020; Ren et al., 2021; Tian et al., 2020; Zhu, 2023). It is important to note that journal tier classifications are reported here as bibliometric indicators of publication venue prestige and research field maturity, not as direct indicators of individual study quality. All included studies underwent independent quality assessment procedures as described in the method section.

Publication and Geographic Trends

Temporal distribution

The temporal trajectory of CDA research over the past decade shows a distinctive growth pattern characterized by three distinct developmental phases. **Figure 2** shows publication increases from 2015 through 2021, with research output expanding from 2 articles in 2015 to a peak of 12 articles in 2021. This growth trajectory subsequently stabilized at approximately 10 articles annually during 2022-2023, suggesting the field has entered a consolidation phase.

Geographic distribution

The geographical distribution of CDA research shows concentration patterns with implications for the global development of this field.

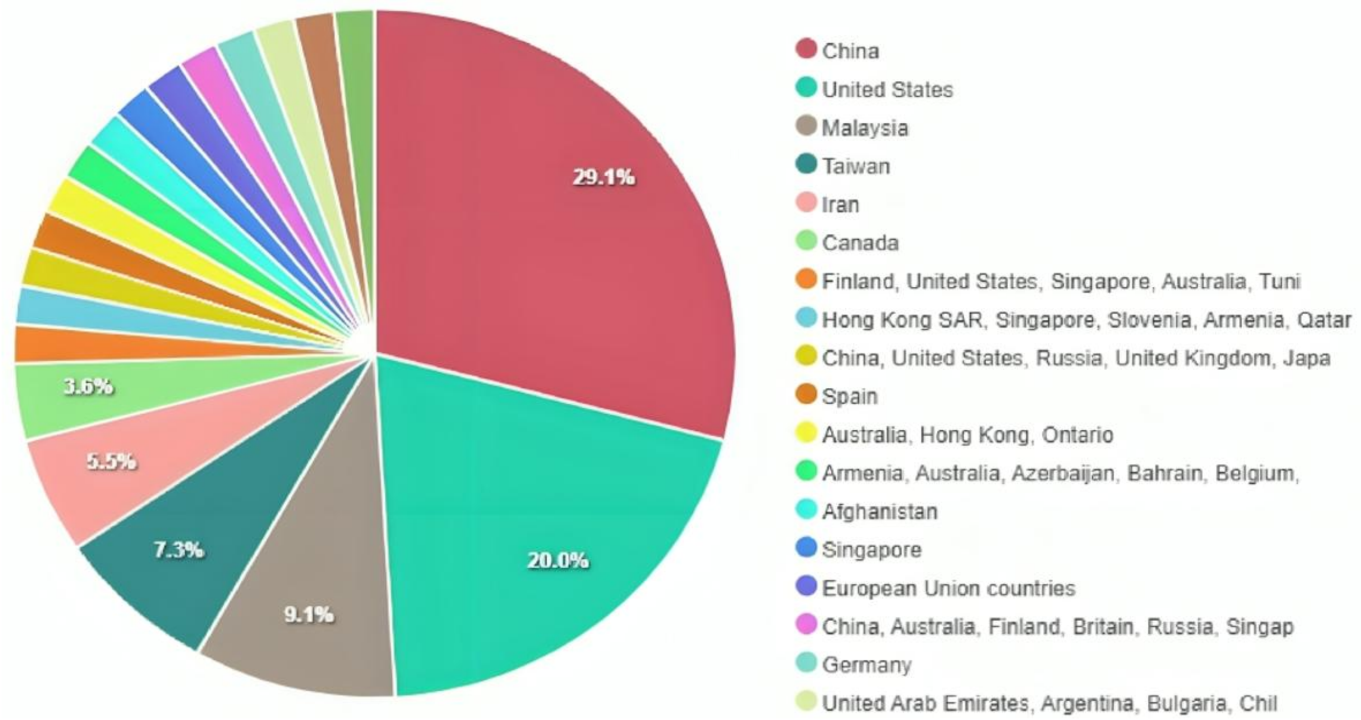

Figure 3. Geographical distribution (Source: Authors' own elaboration, using Watase Uake Tools)

Figure 3 shows that China accounts for 29.1% of all studies, followed by the USA (20%) and Malaysia (9.1%). Taiwan (7.3%) and Iran (5.5%) also emerge as substantial contributors, while other countries including Canada, Singapore, Spain, Afghanistan, and Brazil contribute smaller proportions.

Publication volume, however, does not necessarily correlate directly with academic influence. **Table 2** presents the most-cited studies grouped by country, showing more nuanced dynamics between research productivity and scholarly impact. Although China has produced the highest number of articles (16 studies), the USA actually leads in academic influence with 268 citations across 12 articles. The citation-to-article ratio for the USA reaches 22.3, compared to China's 8.8, indicating that early USA contributions played a foundational role in shaping theoretical and methodological discourse (Leighton & Chu, 2016). This pattern confirms that in emerging scientific fields, the timing and conceptual depth of contributions often prove more decisive for long-term influence than mere publication volume.

TCM Framework Analysis

Theory dimension

Analysis of the 56 studies shows that CDM as the primary theoretical framework, with 28 publications (50%) and a total of 467 citations, reflecting CDM's central position in contemporary CDA research. Other theoretical frameworks shows dispersed distribution patterns, with each theory being employed in only 1-4 studies as illustrated in **Table 3**.

Table 3. Most citation by country

Author	Year	Tier	C	Country	TA
Chen et al., Köhn et al., Le et al., Lei et al., Lin et al., Liu et al., Ma et al., Madison et al., Park et al., Paulsen et al., Skaggs et al., Wang et al.	2014-2025	Q1, Q2	268	USA	12
Dong et al., Hu et al., Huang et al., Jiang et al., Kang et al., Li et al., Mei et al., Meng et al., Qin et al., Ren et al., Tao et al., Tian et al., Tu et al., Wang et al., Wu et al., Wang et al.	2018-2024	Q1, Q2	141	China	16
Mirzaei et al., Ranjbaran et al., Ravand et al.	2017, 2018, 2020	Q1	101	Iran	3
Hung et al., Kuo et al., Shih et al., Wu et al.	2016, 2018	Q1	61	Taiwan	4
Wu et al.	2020	Q1	38	China, USA, Russia, UK, Japan	1
Aryadoust et al.	2018	Q1	32	Singapore	1
Chin et al.	2020-2022	Q1, Q2	28	Malaysia	5
Yamaguchi et al.	2018	Q1	27	Hong Kong, Singapore, Slovenia, Armenia, Qatar	1

Table 3 (Continued). Most citation by country

Author	Year	Tier	C	Country	TA
Cui et al., Leighton et al.	2015	Q1	25	Canada	2
Nájera et al.	2019	Q1	24	Spain	1
Abdulaal et al.	2022	Q1	17	Afghanistan	1
de la Torre et al.	2018	Q1	14	Brazil	1
Wu et al.	2021	Q2	14	China, Australia, Finland, UK, Russia, Singapore	1
Toprak et al.	2021	Q1	6	EU countries	1
Jia et al.	2021	Q1	6	UAE, Argentina, Bulgaria, Chili	1
Maas et al.	2022	Q2	5	Netherland	1
Zhou et al.	2022	Q1	3	Australia, Hong Kong, Canada	1
Delafontaine et al.	2022	Q1	2	Finland, USA, Singapore, Australia, Tunisia	1
Zhu et al.	2023	Q1	1	Armenia, Australia, Azerbaijan, Bahrain, Belgium	1
Wedel et al.	2022	Q1	1	Germany	1

Note. C: Citation & TA: Total articles

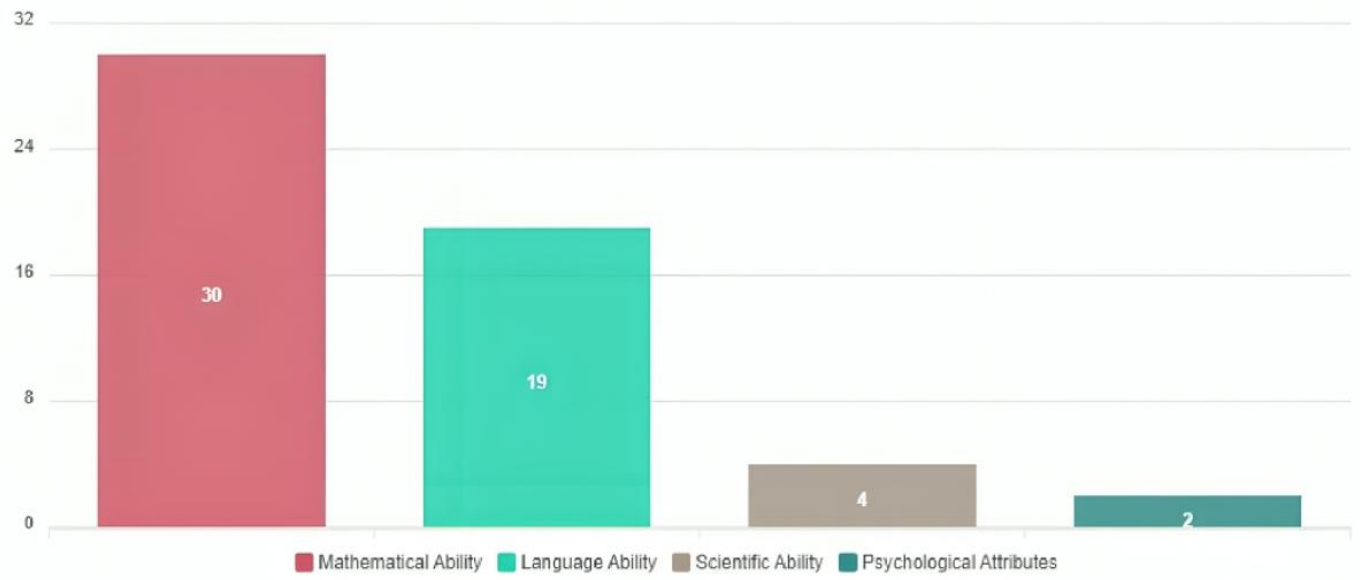

Figure 4. Distribution of subject domains (Source: Authors' own elaboration, using Watase Uake Tools)

Context dimension

Analysis of the 56 studies shows clear concentration on mathematics with 30 studies (Chin & Chew, 2022; Chin et al., 2021a; Wang & Qiu, 2019; Wu et al., 2023), followed by language domains with 19 studies (Aryadoust, 2021; Mei & Chen, 2022), reflecting CDA research focus that aligns with global education policy priorities toward literacy and numeracy (Cui et al., 2016; Lin et al., 2020; Maas et al., 2022; Meng et al., 2023; Nájera et al., 2019).

Science domains demonstrate emerging diversification, though their representation remains limited compared to mathematics and language. Distribution of subject domains shows significant fragmentation, with each non-STEM domain having minimal representation in contemporary CDA research, as illustrated at **Figure 4**.

Method dimension

Analysis of the 56 studies shows that quantitative approaches were employed in 43 studies, while mixed methods and qualitative approaches were used in 7 and 5 studies, respectively, as illustrated in **Figure 5**. This quantitative dominance reflects the methodological orientation of CDA research, which remains deeply rooted in psychometric traditions where statistical precision and empirical validation constitute primary considerations in developing cognitive diagnostic instruments.

Consistent with the quantitative dominance, the distribution of analytical methods shows clear concentration on specific CDMs. DINA model and G-DINA each lead with 6 studies, followed by attribute hierarchy method (Köhn & Chiu, 2019; Tu et al., 2019) and combinations of multiple CDMs (DINA, DINO,

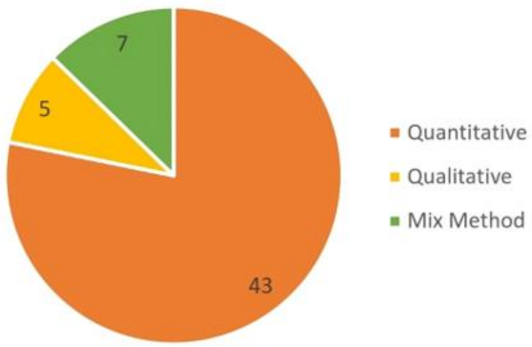

Figure 5. Method used (Source: Authors' own elaboration, using Watase Uake Tools)

RRUM, LLM, ACDM, GDM, LCDM, and GDINA), each employed in 3 studies. Other analytical methods show more dispersed diversification, including maximum likelihood estimation (MLE) and ANOVA each appearing in 2 studies, along with various innovative approaches such as neural network parameter optimization, machine learning, discrimination indices, and Boolean operations, each utilized in single studies (**Figure 6**). This distribution indicates methodological evolution in CDA research from traditional approaches toward integration with more advanced computational technologies.

Research Focus Distribution

The distribution of research aspects in CDA shows a clear dominance of ability classification, represented by 29 studies, followed by Q-matrix validation (11 studies), learning pathway analysis (9 studies), and educational intervention effectiveness (6 studies) (**Figure 7**). This distribution underscores a substantial imbalance, as diagnostic aspects (ability classification) account for nearly half of the reviewed studies, whereas applied and intervention-oriented dimensions receive considerably less attention.

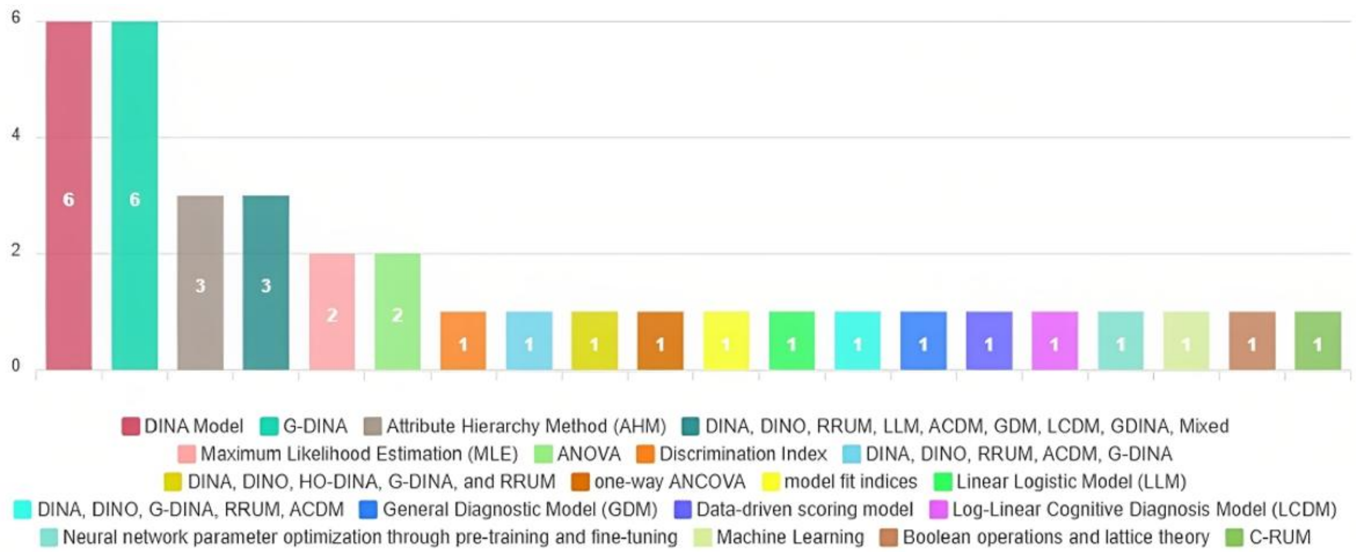

Figure 6. Distribution of analytical methods (Source: Authors' own elaboration, using Watase Uake Tools)

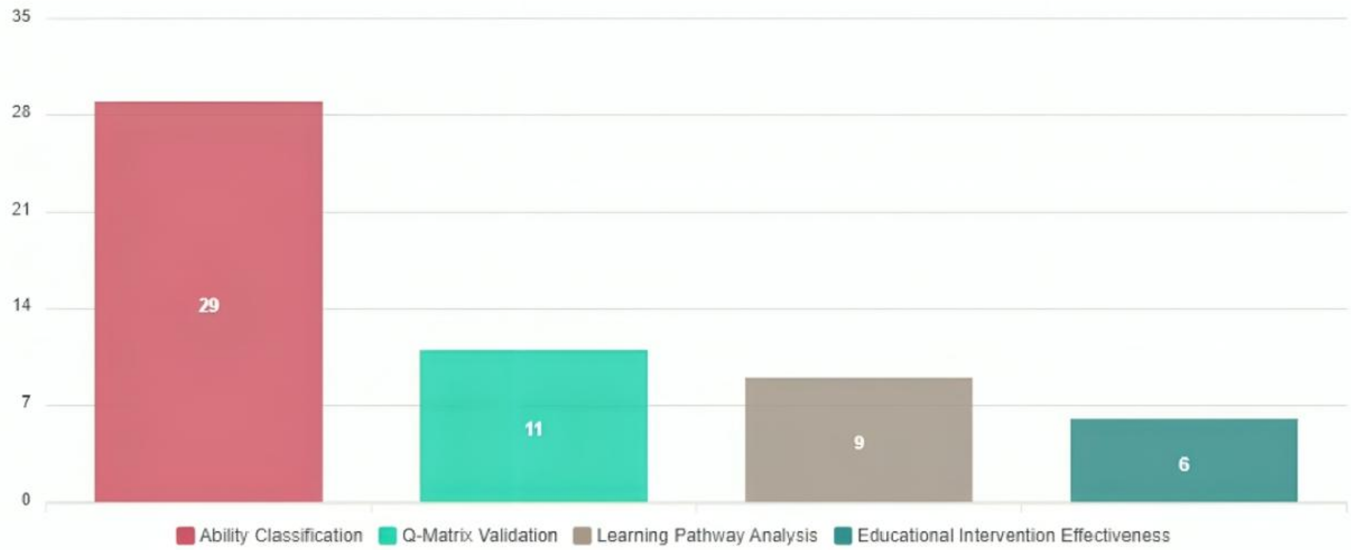

Figure 7. Distribution of research aspects (Source: Authors' own elaboration, using Watase Uake Tools)

Table 4. Temporal distribution of research aspects

Authors	Year	Research aspect
Aryadoust et al., Chin et al., Cui et al., Dong et al., Hu et al., Hung et al., Kuo et al., Le et al., Lei et al., Leighton et al., Li et al., Maas et al., Mei et al., Meng et al., Mirzaei et al., Park et al., Paulsen et al., Ranjbaran et al., Ravand et al., Shih et al., Skaggs et al., Tao et al., Wang et al., Wu et al., Yamaguchi et al.	2015-2025	Ability classification
Da et al., Delafontaine et al., Kang et al., Kohn et al., Liu et al., Ma et al., Madison et al., Nájera et al., Qin and Guo, Tian, Tu et al., Wang et al.	2015, 2016, 2018, 2019, 2020, 2021, 2022, 2023	Q-matrix validation
Chen et al., Jia et al., Jiang et al., Lin et al., Toprak et al., Wu et al., Zhou et al., Zhu et al.	2020-2024	Learning pathway analysis
Abdulaal et al., Chin et al., Huang et al., Ren et al., Wedel et al., Wu et al.	2018, 2020, 2021, 2022	Educational intervention effectiveness

A temporal analysis further indicates divergent developmental trajectories across research aspects. Ability classification has maintained consistent continuity from 2015 to 2025, involving the most diverse range of contributors. Q-matrix validation shows steady growth between 2015 and 2023, while learning pathway analysis emerges as a new trend gaining momentum after 2020. In contrast, research on intervention effectiveness remains the most limited (Abdulaal et al., 2022), with scattered representation only between 2018 and 2022 (**Table 4**).

Collectively, these patterns suggest that although CDA's methodological foundations have been consolidated, the integration of diagnostic insights into instructional practice remains underdeveloped.

DISCUSSION

This SLR synthesized 56 studies on CDA and CDMs published 2015-2025, addressing four fundamental research questions regarding CDA/CDM evolution, theoretical and methodological contributions, applications across educational contexts, and existing research gaps. The discussion integrates findings from multiple analytical dimensions to provide comprehensive understanding of the field's current state and future trajectories.

Evolution of CDA and CDMs in Education Over the Past Decade (2015-2025)

The temporal analysis shows that CDA/CDM research has undergone three distinct developmental phases over the past decade, each characterized by unique theoretical emphases and methodological sophistication. The foundational period (2015-2017) concentrated primarily on core methodological refinements, particularly Q-matrix validation and diagnostic model specification (Lei & Li, 2016). This foundational work established theoretical groundwork essential for subsequent applications, with studies focusing on establishing psychometric properties and validating basic model assumptions.

The transitional period (2018-2020) witnessed a paradigmatic shift toward empirical validation through classroom applications and cross-national comparative studies (Wu et al., 2020; Yamaguchi & Okada, 2018). Researchers began examining the practical utility of diagnostic models across diverse educational contexts, moving beyond purely theoretical investigations. The year 2018 alone generated 182 citations, reflecting the substantial scholarly impact of foundational works that established methodological standards and theoretical frameworks during this period.

Since 2021, the field has entered what might be characterized as an integration phase, marked by notable shifts toward incorporating CDA with emerging technologies. The observed surge through 2021 coincides with several convergent developments in educational assessment and technology. The proliferation of machine learning and artificial intelligence applications in psychometrics created novel methodological possibilities for CDA implementation (Cui et al., 2016; Qin & Guo, 2023). Concurrently, the COVID-19 pandemic accelerated adoption of digital assessment platforms, generating unprecedented demand for diagnostic tools capable of functioning effectively in online learning environments. This convergence provided both technological infrastructure and practical urgency that propelled CDA research forward.

Contemporary research emphasizes development of adaptive testing systems, neural network-enhanced diagnostic models, and real-time feedback mechanisms integrated with online learning platforms (Tao et al., 2024). This technological integration represents more than simple digitization; it signifies CDA's evolution from a specialized psychometric tool into a component of comprehensive, technology-driven educational ecosystems. The stabilization of publication volume after 2021, rather than indicating declining interest, suggests the field is consolidating gains while emphasizing quality over quantity. This pattern typically emerges when research domains transition from rapid exploration to systematic application and refinement.

However, this evolution has occurred primarily within psychometric paradigms. The pronounced concentration on specific CDMs may also create methodological homogeneity that potentially constrains innovation and theoretical advancement. The dominance of particular analytical approaches might reflect not only their effectiveness but also inertial tendencies within academic communities, where established methods tend to be perpetuated without sufficient exploration of alternative approaches.

Theoretical and Methodological Contributions in CDA/CDM Research

The review identifies substantial theoretical and methodological contributions while revealing critical areas requiring further development. The dominance of CDM as the primary theoretical framework, with 28 publications (50%) and 467 citations, reflects CDM's central position in contemporary CDA research. The pronounced concentration on CDM, particularly the DINA and G-DINA models, confirms their foundational position and reflects methodological maturity and standardization within CDA research, yet simultaneously shows constraints in theoretical diversification (Leighton & Chu, 2016; Ma & de la Torre, 2020).

The elevated citation ratio for CDM compared to other theoretical frameworks shows not merely popularity but also substantial academic impact, suggesting that CDM-based research tends to generate more influential contributions within the scholarly community (Hung & Huang, 2019; Shih et al., 2019). The substantial concentration of publications within Q1 journals (49 studies, 87.5%) signals the methodological sophistication of CDA research, which has achieved standards recognized by high caliber journals, particularly within psychology and educational measurement domains (Lin et al., 2020; Ren et al., 2021).

Methodologically, the overwhelming quantitative dominance (43 studies, 76.8%) confirms CDA's position as a discipline firmly anchored within classical psychometric traditions, where statistical validation and measurement precision represent core concerns (de la Torre, 2011; Templin & Henson, 2010). This methodological preference shows the research community's commitment to scientific rigor and empirical validation, consistent with positivist paradigms that have historically shaped educational measurement. The concentration on quantitative methods also indicates the development of sophisticated methodological toolkits designed to address technical complexities in diagnostic modeling, reflecting CDA's maturation as a research domain.

The emergence of hybrid approaches, such as Q-matrix integration with neural networks and CDM combination with latent growth curve modeling (Kuo et al., 2016; Park et al., 2018), marks a transitional phase in CDA research toward more integrative and multidimensional paradigms (Da Silva et al., 2019; Qin & Guo, 2023; Tao et al., 2024). While these innovations enhance accuracy and flexibility, the highly dispersed distribution of alternative theoretical frameworks (each appearing in only 1-2 studies) shows fragmentation in theoretical exploration and limited integration with broader learning theories such as constructivism or sociocultural perspectives.

However, the markedly limited utilization of qualitative approaches (8.9%) and mixed methods (12.5%) shows a fundamental gap in understanding CDA's practical implementation. The scarcity of qualitative perspectives means CDA research remains predominantly focused on technical precision without adequately exploring how diagnostic results are interpreted, utilized, and experienced within authentic educational contexts (Paulsen & Valdivia, 2022). This gap creates a disconnect between statistical sophistication and practical utility, where highly accurate diagnostic models may possess limited applicability if not accompanied by understanding of user needs, contextual constraints, and implementation challenges.

The constrained presence of Vygotsky's sociocultural theory (1 study, 17 citations) and theory of learning from error (1 study, 4 citations) indicates a significant gap between diagnostic assessment practices and contemporary learning theories, despite the potential of such integration to enrich understanding of cognitive processes underlying student performance. These findings suggest that while CDA research has achieved considerable methodological sophistication, substantial opportunities exist for more holistic theoretical development through cross-disciplinary engagement (Kang et al., 2019; Madison & Bradshaw, 2015). The extreme dominance of CDM (50% of all studies) may create academic silos that constrain theoretical innovation and practical applications of CDA within broader learning contexts.

Applications of CDA/CDMs Across Educational Contexts

The geographic and contextual analysis shows both breadth and notable concentrations in CDA/CDM applications. The pronounced Asian dominance in CDA research contributions, with China leading at 29.1% of all studies, followed by the USA (20%) and Malaysia (9.1%), reflects substantial investment in educational research infrastructure and policy priorities positioning diagnostic assessment as key components in educational system enhancement. This concentration aligns with regional emphases on mathematics and science education excellence and responses to international assessment pressures from PISA and TIMSS (Delafontaine et al., 2022; Wu et al., 2020).

Publication volume, however, does not necessarily correlate directly with academic influence. Although China has produced the highest number of articles (16 studies), the USA actually leads in academic influence with 268 citations across 12 articles. The citation-to-article ratio for the USA reaches 22.3, compared to China's 8.8, indicating that early U.S. contributions played a foundational role in shaping theoretical and methodological discourse (Leighton & Chu, 2016). This pattern confirms that in emerging scientific fields, the timing and conceptual depth of contributions often prove more decisive for long-term influence than mere publication volume.

Citation patterns also show complex knowledge transfer dynamics between Western and Asian contexts. Ma and de la Torre's (2020) contributions from the USA remain widely cited, particularly for their methodological innovations in Q-matrix validation, which have subsequently been adopted and adapted in Asian studies (Ren et al., 2021; Tian et al., 2020). This interplay illustrates a dynamic process of knowledge transfer wherein early Western frameworks have been adapted, refined, and localized within Asian contexts, creating methodological evolution responsive to regional needs.

The strong regional concentration in Asia suggests that CDA has become increasingly localized to meet national reform agendas while simultaneously responding to global performance pressures. This dominance reflects substantial investment in educational research infrastructure and policy priorities. Nevertheless, this pattern raises critical questions regarding the generalizability of findings. International comparative studies become essential to uncover how contextual factors, including curriculum design, classroom culture, and teacher preparation, shape the diagnostic validity of CDA (Delafontaine et al., 2022; Wu et al., 2022).

Subject domain analysis shows the overwhelming concentration on mathematics (30 studies, 53.6%) and language (19 studies, 33.9%), creating construct underrepresentation that threatens validity inferences regarding broader educational competencies (Messick & Linn, 1989). While this dominance can be understood through accessibility theory because both domains possess readily interpretable construct representations, it overlooks the developmental potential of CDA within science domains that shows promising structural characteristics. Science domains, particularly physics and chemistry, possess conceptual hierarchies that can be decomposed into specific cognitive attributes, similar to mathematics but with heightened applicative complexity (Le et al., 2025).

Research within science domains shows encouraging developments, where CDA can identify specific misconceptions and patterns of scientific reasoning among students (Chen et al., 2025; Zhou & Traynor, 2022). Unlike domains such as creativity that encounter "measurement paradox" where the most meaningful constructs often prove least measurable (Borsboom, 2006), science domains offer balance between cognitive complexity and measurability that can bridge the gap between structured domains (mathematics) and other contextual domains. Expansion into science domains also aligns with 21st century skills requirements that emphasize scientific literacy and critical thinking capabilities (Pellegrino et al., 2001).

CDA development within science domains can open pathways for interdisciplinary applications that integrate mathematical reasoning, scientific inquiry, and language skills in unified approaches (Hu et al., 2021). This would address limitations in understanding knowledge transfer across contexts, where meaningful learning occurs when students can apply knowledge flexibly (Bransford et al., 2000). Science domains as bridges toward broader diversification can facilitate development of diagnostic approaches that accommodate authentic learning environments characterized by inherent interdisciplinarity while maintaining methodological rigor.

Research Gaps and Limitations Requiring Future Attention

The synthesis shows several critical gaps requiring scholarly attention to advance CDA/CDM research and practice. The distribution of research shows a clear dominance of ability classification, represented by 29 studies (51.8%), while educational intervention effectiveness receives considerably less attention with only 6 studies (10.7%). This distribution underscores a substantial imbalance, as diagnostic aspects account for nearly half of the reviewed studies, whereas applied and intervention-oriented dimensions receive considerably less attention.

The dominance of ability classification resonates with classical cognitive assessment paradigms, which emphasize the identification of learner profiles (Leighton & Gierl, 2007; Pellegrino et al., 2001). However, this overemphasis on diagnosis without subsequent pedagogical action reinforces critiques of the "assessment of learning" paradigm (Black & Wiliam, 1998). Conversely, the underrepresentation of studies addressing intervention effectiveness suggests that feedback confined to diagnosis, without the support of targeted instructional strategies, contributes little to meaningful gains in student achievement (Abdulaal et al., 2022; Hattie & Timperley, 2007). Such imbalance runs counter to the foundational principles of formative assessment theory (Bloom, 1971), which emphasize that the true value of assessment lies in its ability to inform and improve learning processes rather than merely identify cognitive deficits.

The paucity of intervention-focused research contrasts with the principles of design-based research (Brown, 1992; Collins, 1992), which stresses iterative cycles of diagnosis, intervention, and evaluation. This gap raises concerns regarding consequential validity, since assessments that fail to demonstrably improve learning outcomes have limited educational value (Messick & Linn, 1989). The imbalance between CDA's technical sophistication and its practical application in authentic classrooms points to what can be termed an ecological gap (Wedel et al., 2022). Transfer theory further reinforces this concern: diagnostic insights generated in assessment contexts do not automatically translate into instructional improvement without explicit bridging mechanisms (Barnett & Ceci, 2002; Skaggs et al., 2016).

The fragmentation of alternative theoretical frameworks shows the need for consolidation and synthesis to develop more comprehensive theoretical frameworks that can integrate CDM's methodological strengths with insights from learning theory, cognitive psychology, and educational technology to create assessment approaches more responsive to the complexities of contemporary learning processes. The limited representation from African, South American, and Middle Eastern countries suggests untapped potential for expanding CDA research to encompass broader educational environments, diverse curriculum designs, and varied student populations.

Methodologically, the gap between technical sophistication and practical relevance requires urgent attention. The limited integration with qualitative insights means the research community potentially overlooks critical understanding of how diagnostic information translates into meaningful educational interventions. Future research must address this methodological imbalance through more systematic integration of mixed-methods approaches that can capture both statistical precision and contextual richness, enabling development of diagnostic tools that are not only technically robust but also practically meaningful and implementable across diverse educational settings.

The limited diversification beyond core academic subjects represents missed opportunities for demonstrating CDA's broader applicability. The concentration on traditional academic subjects may reflect measurement convenience rather than educational priority, suggesting that future research should strategically expand into domains addressing evolving educational needs and workforce requirements. The fragmentation in innovative method usage suggests early-stage exploration that has not yet achieved critical mass for substantial impact on field development, indicating that technological integration remains in exploration stages requiring systematic validation.

Limitations

This systematic review has several methodological limitations requiring acknowledgment.

First, the exclusive reliance on Scopus database and restriction to English-language publications may have excluded relevant studies from other databases and non-English contexts, potentially introducing geographic and linguistic biases.

Second, 54 articles were inaccessible despite retrieval efforts, raising concerns about potential selection bias if excluded studies systematically differed from included ones.

Third, while the quality assessment achieved acceptable inter-rater reliability ($\kappa = 0.80$), the domain-specific checklist and inclusion threshold (≥7/10) involved subjective judgment and lacked external validation.

Fourth, the TCM framework classification, despite high inter-coder agreement (92.3%), required interpretive decisions for studies with multiple theoretical frameworks or hybrid methodologies, potentially oversimplifying methodological nuances.

Fifth, the review may be subject to publication bias favoring significant findings and rigorous methodologies, potentially underrepresenting null findings or unsuccessful CDA/CDM implementations. The geographic concentration in Asia and North America limits generalizability to other educational contexts.

Finally, the narrative synthesis approach does not provide quantitative effect size estimates that meta-analysis would offer, and interpretative claims reflect the authors' theoretical perspectives. Despite these limitations, transparent reporting procedures, documented inter-rater reliability, and comprehensive appendices enable critical evaluation of the review's contributions to CDA/CDM research.

CONCLUSION

This SLR examined 56 studies on CDA and CDMs published 2015-2025. Using the TCM framework as the organizing lens, the review shows that CDA/CDMs have evolved from classical psychometric models such as DINA and G-DINA to more advanced approaches integrating machine learning, neural networks, and adaptive platforms. While quantitative designs dominate, the limited use of qualitative and mixed-methods studies highlights a persistent gap in understanding how diagnostic results are implemented in practice.

From a contextual perspective, the findings indicate that CDA/CDMs are most widely applied in mathematics, science, and language education, with strong regional growth in East Asia and continuing theoretical contributions from the USA. From a theoretical perspective, advances are evident in hybrid diagnostic models, yet overreliance on CDM paradigms suggests the need for broader engagement with contemporary learning theories. From a methodological perspective, the field is marked by technical sophistication yet hampered by weak integration with classroom practice.

Practically, CDA/CDMs have enhanced feedback precision, supported curriculum reforms through large-scale assessments, and offered potential for personalized learning. However, the imbalance between technical sophistication and instructional application underscores the importance of bridging diagnostic insights with pedagogical action. Future research should expand geographically, adopt longitudinal designs, integrate interdisciplinary perspectives, and engage teachers in tool development. Overall, CDA/CDMs should be reframed not merely as technical instruments but as transformative frameworks to advance equity, personalization, and evidence-based policy in global education.

Author contributions: FH: conceptualization, theoretical framework development, methodology, data collection, data analysis, visualization, writing – original draft; **AH:** conceptualization, research design, supervision, validation, writing – review & editing; **RA:** methodology, data curation, formal analysis, writing – review & editing; **BY:** supervision, validation, critical review of the manuscript, writing – review & editing; M: supervision, validation, critical revision of the manuscript, writing – review & editing.All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study is based on an analysis of published research articles and does not involve human participants, human biological materials, or personal/sensitive data. Therefore, ethical approval from an institutional review board or ethics committee was not required. All data sources used in this study are publicly available and properly cited in accordance with ethical standards of academic research.

AI statement: The authors stated that generative artificial intelligence (AI) tools were used in a limited manner to assist with language editing and clarity of expression during manuscript preparation. The use of AI tools did not affect the study design, data analysis, interpretation of results, or conclusions. The authors take full responsibility for the content of this article.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Abdulaal, M. A. A.-D., Alenazi, M. H., Tajuddin, A. J. A., & Hamidi, B. (2022). Dynamic vs. diagnostic assessment: Impacts on EFL learners' speaking fluency and accuracy, learning anxiety, and cognitive load. *Language Testing in Asia, 12*(1), Article 32. https://doi.org/10.1186/s40468-022-00179-0

Aryadoust, V. (2021). A Cognitive diagnostic assessment study of the listening test of the Singapore-Cambridge general certificate of education o-level: Application of DINA, DINO, G-DINA, HO-DINA, and RRUM. *International Journal of Listening, 35*(1), 29-52. https://doi.org/10.1080/10904018.2018.1500915

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), Article 612. https://doi.org/10.1037/0033-2909.128.4.612

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Bloom, B. S. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*(3), 425-440. https://doi.org/10.1007/s11336-006-1447-6

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn (vol. 11)*. National Academy Press.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141-178. https://doi.org/10.1207/s15327809jls0202_2

Chen, Y.-J. I., Wu, Y.-J., Chen, Y.-H., & Irey, R. (2025). Development and initial validation of the computer-based orthographic processing assessment short form: An application of cognitive diagnostic modeling. *Journal of Psychoeducational Assessment, 43*(3), 310-327. https://doi.org/10.1177/07342829241304165

Chin, H., & Chew, C. M. (2022). Online cognitive diagnostic assessment with ordered multiple-choice items for word problems involving 'time.' *Education and Information Technologies, 27*(6), 7721-7748. https://doi.org/10.1007/s10639-022-10956-2

Chin, H., Chew, C. M., & Lim, H. L. (2021a). Development and validation of online cognitive diagnostic assessment with ordered multiple-choice items for 'multiplication of time.' *Journal of Computers in Education, 8*(2), 289-316. https://doi.org/10.1007/s40692-020-00180-7

Chin, H., Chew, C. M., & Lim, H. L. (2021b). Incorporating feedback in online cognitive diagnostic assessment for enhancing grade five students' achievement in 'time.' *Journal of Computers in Education, 8*(2), 183-212. https://doi.org/10.1007/s40692-020-00176-3

Collins, A. (1992). Toward a design science of education. In E. Scanlon, & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). Springer. https://doi.org/10.1007/978-3-642-77750-9_2

Cui, Y., Gierl, M., & Guo, Q. (2016). Statistical classification for cognitive diagnostic assessment: An artificial neural network approach. *Educational Psychology, 36*(6), 1065-1082. https://doi.org/10.1080/01443410.2015.1062078

Da Silva, M. A., Liu, R., Huggins-Manley, A. C., & Bazán, J. L. (2019). Incorporating the Q-matrix into multidimensional item response theory models. *Educational and Psychological Measurement, 79*(4), 665-687. https://doi.org/10.1177/0013164418814898

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179-199. https://doi.org/10.1007/s11336-011-9207-7

Delafontaine, J., Chen, C., Park, J. Y., & Van Den Noortgate, W. (2022). Using country-specific Q-matrices for cognitive diagnostic assessments with international large-scale data. *Large-Scale Assessments in Education, 10*(1), Article 19. https://doi.org/10.1186/s40536-022-00138-4

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

Hu, T., Yang, J., Wu, R., & Wu, X. (2021). An international comparative study of students' scientific explanation based on cognitive diagnostic assessment. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.795497

Huang, R., Liu, Z., Zi, D., Huang, Q., & Pan, S. (2022). A multi-level remedial teaching design based on cognitive diagnostic assessment: Taking the electromagnetic induction as an example. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.851378

Hung, S.-P., & Huang, H.-Y. (2019). A sequential process model for cognitive diagnostic assessment with repeated attempts. *Applied Psychological Measurement, 43*(7), 495-511. https://doi.org/10.1177/0146621618813111

Jia, B., Zhu, Z., & Gao, H. (2021). International comparative study of statistics learning trajectories based on PISA data on cognitive diagnostic models. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.657858

Jiang, B., Li, X., Yang, S., Kong, Y., Cheng, W., Hao, C., & Lin, Q. (2022). Data-driven personalized learning path planning based on cognitive diagnostic assessments in MOOCs. *Applied Sciences, 12*(8), 3982. https://doi.org/10.3390/app12083982

Kang, C., Yang, Y., & Zeng, P. (2019). Q-matrix refinement based on item fit statistic RMSEA. *Applied Psychological Measurement, 43*(7), 527-542. https://doi.org/10.1177/0146621618813104

Köhn, H.-F., & Chiu, C.-Y. (2019). Attribute hierarchy models in cognitive diagnosis: Identifiability of the latent attribute space and conditions for completeness of the Q-matrix. *Journal of Classification, 36*(3), 541-565. https://doi.org/10.1007/s00357-018-9278-6

Kuo, B.-C., Chen, C.-H., Yang, C.-W., & Mok, M. M. C. (2016). Cognitive diagnostic models for tests with multiple-choice and constructed-response items. *Educational Psychology, 36*(6), 1115-1133. https://doi.org/10.1080/01443410.2016.1166176

Le, V., Nissen, J. M., Tang, X., Zhang, Y., Mehrabi, A., Morphew, J. W., Chang, H. H., & Van Dusen, B. (2025). Applying cognitive diagnostic models to mechanics concept inventories. *Physical Review Physics Education Research, 21*, Article 010103. https://doi.org/10.1103/PhysRevPhysEducRes.21.010103

Leighton, J. P., & Chu, M.-W. (2016). First among equals: Hybridization of cognitive diagnostic assessment and evidence-centered game design. *International Journal of Testing, 16*(2), 164-180. https://doi.org/10.1080/15305058.2015.1107075

Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611186

Li, Y., Zhen, M., & Liu, J. (2021). Validating a reading assessment within the cognitive diagnostic assessment framework: Q-matrix construction and model comparisons for different primary grades. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.786612

Lin, Q., Xing, K., & Park, Y. S. (2020). Measuring skill growth and evaluating change: Unconditional and conditional approaches to latent growth cognitive diagnostic models. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.02205

Ma, W., & de la Torre, J. (2020). An empirical Q-matrix validation method for the sequential generalized DINA model. *British Journal of Mathematical and Statistical Psychology, 73*(1), 142-163. https://doi.org/10.1111/bmsp.12156

Maas, L., Brinkhuis, M. J. S., Kester, L., & Wijngaards-de Meij, L. (2022). Cognitive diagnostic assessment in university statistics education: Valid and reliable skill measurement for actionable feedback using learning dashboards. *Applied Sciences, 12*(10), Article 4809. https://doi.org/10.3390/app12104809

Madison, M. J., & Bradshaw, L. P. (2015). The effects of Q-matrix design on classification accuracy in the log-linear cognitive diagnosis model. *Educational and Psychological Measurement, 75*(3), 491-511. https://doi.org/10.1177/0013164414539162

Mei, H., & Chen, H. (2022). Assessing students' translation competence: Integrating China's standards of English with cognitive diagnostic assessment approaches. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.872025

Meng, Y., Wang, Y., & Zhao, N. (2023). Cognitive diagnostic assessment of EFL learners' listening barriers through incorrect responses. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1126106

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Macmillan Publishing Co, Inc; American Council on Education.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339*, Article b2535. https://doi.org/10.1136/bmj.b2535

Nájera, P., Sorrel, M. A., & Abad, F. J. (2019). Reconsidering cutoff points in the general method of empirical Q-matrix validation. *Educational and Psychological Measurement, 79*(4), 727-753. https://doi.org/10.1177/0013164418822700

Panic, N., Leoncini, E., de Belvis, G., Ricciardi, W., & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS ONE, 8*(12), Article e83138. https://doi.org/10.1371/journal.pone.0083138

Park, Y. S., Xing, K., & Lee, Y.-S. (2018). Explanatory cognitive diagnostic models: Incorporating latent and observed predictors. *Applied Psychological Measurement, 42*(5), 376-392. https://doi.org/10.1177/0146621617738012

Paulsen, J., & Valdivia, D. S. (2022). Examining cognitive diagnostic modeling in classroom assessment conditions. *The Journal of Experimental Education, 90*(4), 916-933. https://doi.org/10.1080/00220973.2021.1891008

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing, learning, and instruction*. National Academy Press.

Qin, H., & Guo, L. (2023). Using machine learning to improve Q-matrix validation. *Behavior Research Methods, 56*(3), 1916-1935. https://doi.org/10.3758/s13428-023-02126-0

Ranjbaran, F., & Alavi, S. M. (2017). Developing a reading comprehension test for cognitive diagnostic assessment: A RUM analysis. *Studies in Educational Evaluation, 55*, 167-179. https://doi.org/10.1016/j.stueduc.2017.10.007

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology, 38*(10), 1255-1277. https://doi.org/10.1080/01443410.2018.1489524

Ren, H., Xu, N., Lin, Y., Zhang, S., & Yang, T. (2021). Remedial teaching and learning from a cognitive diagnostic model perspective: Taking the data distribution characteristics as an example. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.628607

Shih, S.-C., Kuo, B.-C., & Lee, S.-J. (2019). An online game-based computational estimation assessment combining cognitive diagnostic model and strategy analysis. *Educational Psychology, 39*(10), 1255-1277. https://doi.org/10.1080/01443410.2018.1501468

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70*, 747-770. https://doi.org/10.1146/annurev-psych-010418-102803

Skaggs, G., Hein, S. F., & Wilkins, J. L. M. (2016). Diagnostic profiles: A standard setting method for use with a cognitive diagnostic model. *Journal of Educational Measurement, 53*(4), 448-458. https://doi.org/10.1111/jedm.12125

Tao, J., Zhao, W., Liu, F., Guo, X., Cheng, N., Guo, Q., Xu, X., & Duan, H. (2024). Cognitive diagnosis method via q-matrix-embedded neural networks. *Applied Sciences, 14*(22), Article 10380. https://doi.org/10.3390/app142210380

Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Routledge. https://doi.org/10.4324/9780203883372

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), Article 287. https://doi.org/10.1037/1082-989X.11.3.287

Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Tian, W., Zhang, J., Peng, Q., & Yang, X. (2020). Q-matrix designs of longitudinal diagnostic classification models with hierarchical attributes for formative assessment. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.01694

Toprak-Yildiz, T. E. (2021). An international comparison using cognitive diagnostic assessment: Fourth graders' diagnostic profile of reading skills on PIRLS 2016. *Studies in Educational Evaluation, 70*, Article 101057. https://doi.org/10.1016/j.stueduc.2021.101057

Tu, D., Wang, S., Cai, Y., Douglas, J., & Chang, H.-H. (2019). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted Q-matrix design. *Applied Psychological Measurement,* *43*(4), 255-271. https://doi.org/10.1177/0146621618765721

Wang, W.-C., & Qiu, X.-L. (2019). Multilevel modeling of cognitive diagnostic assessment: The multilevel DINA example. *Applied Psychological Measurement, 43*(1), 34-50. https://doi.org/10.1177/0146621618765713

Wedel, A., Müller, C. R., & Greiner, F. (2022). Diagnostic cases in pre-service teacher education: Effects of text characteristics and empathy on text-based cognitive models. *Educational Psychology, 42*(6), 694-713. https://doi.org/10.1080/01443410.2022.2047615

Wu, H.-M. (2019). Online individualised tutor for improving mathematics learning: A cognitive diagnostic model approach. *Educational Psychology, 39*(10), 1218-1232. https://doi.org/10.1080/01443410.2018.1494819

Wu, X., Wu, R., Chang, H.-H., Kong, Q., & Zhang, Y. (2020). International comparative study on PISA mathematics achievement test based on cognitive diagnostic models. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.02230

Wu, X., Xu, T., & Zhang, Y. (2023). Research on the data analysis knowledge assessment of pre-service teachers from China based on cognitive diagnostic assessment. *Current Psychology, 42*(6), 4885-4899. https://doi.org/10.1007/s12144-021-01836-y

Wu, X., Zhang, Y., Wu, R., & Chang, H.-H. (2022). A comparative study on cognitive diagnostic assessment of mathematical key competencies and learning trajectories: PISA data analysis based on 19, 454 Students from 8 countries. *Current Psychology, 41*(11), 7854-7866. https://doi.org/10.1007/s12144-020-01230-0

Yamaguchi, K., & Okada, K. (2018). Comparison among cognitive diagnostic models for the TIMSS 2007 fourth grade mathematics assessment. *PLoS ONE, 13*(2), Article e0188691. https://doi.org/10.1371/journal.pone.0188691

Zhou, S., & Traynor, A. (2022). Measuring students' learning progressions in energy using cognitive diagnostic models. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.892884

Zhu, Z. (2023). International comparative study of learning trajectories based on TIMSS 2019 G4 data on cognitive diagnostic models. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1241656

APPENDIX A

Table A1. Complete TCM classification matrix

No	Study	Country	Theory (T)	Context (C)	Method (M)	Research focus
1	Le et al. (2025)	USA	CDM	Science, higher ed	Mixed-methods, DINA model	Ability classification
2	Chen et al. (2024)	USA	CDM	Writing	Quantitative, DINA model	Ability classification
3	Tao et al. (2024)	China	CDM	General/methodological	Quantitative, neural network parameter optimization through pre-training and fine-tuning	Q-matrix validation
4	Meng et al. (2023)	China	CDM	Reading, higher ed	Quantitative, CDMs	Ability classification
5	Zhu (2023)	Armenia	CDM	Mathematics	Quantitative, general diagnostic model	Ability classification
6	Qin and Guo (2023)	China	Machine learning	General/methodological	Quantitative, machine learning	Q-matrix validation
7	Wedel et al. (2022)	Germany	Construction-integration theory	General/methodological, teacher Ed	Quantitative, multiple hierarchical regression analysis	Ability classification
8	Chin and Chew (2022)	Malaysia	CDA+NEA	Mathematics, primary	Quantitative, attribute hierarchy method	Ability classification
9	Chin and Chew (2022)	Malaysia	CDA+NEA	General/methodological	Quantitative, hierarchical consistency index	Ability classification
10	Abdulaal et al. (2022)	Afghanistan	Vygotsky's ZPD	Language	Quantitative, ANOVA	Ability classification
11	Delafontaine et al. (2022)	Finland	CDM	General/methodological	Quantitative, G-DINA	Ability classification
12	Huang et al. (2022)	China	Cognitive diagnosis theory	General/methodological	Quantitative, ANOVA	Ability classification
13	Mei and Chen (2022)	China	CDA	Language	Quantitative, LLM	Ability classification
14	Zhou and Traynor (2022)	Australia, Hong Kong, Canada	CDM	Science, primary	Quantitative, CDMs	Ability classification
15	Jiang et al. (2022)	China	Data-driven scoring model	General/methodological	Quantitative, data-driven scoring model	Ability classification
16	Maas et al. (2022)	Netherlands	CDM	General/methodological, higher ed	Quantitative, DCMs	Ability classification
17	Ren et al. (2021)	China	CDM	Mathematics, secondary	Quantitative, DINA Model	Ability classification
18	Dong et al. (2021)	China	CDM	General/methodological	Quantitative, model selection using Wald test	Ability classification
19	Jia et al. (2021)	UAE	CDM	General/methodological	Quantitative, CDMs	Ability classification
20	Wang et al. (2021)	China	MLE	General/methodological	Quantitative, MLE	Ability classification
21	Wu et al., 2021	China	CDM	General/methodological, teacher ed	Quantitative, CDM	Ability Classification
22	Chin et al., 2021	Malaysia	CDA+NEA	General/methodological, primary	Quantitative, attribute hierarchy method	Ability Classification
23	Wu et al., 2021	China	CDM	Mathematics	Quantitative, CDM	Ability Classification
24	Chin et al., 2021	Malaysia	Assessment triangle & BEAR assessment system	Mathematics	Quantitative, attribute hierarchy method	Ability classification
25	Toprak-Yildiz, 2021	EU countries	CDA	Reading	Quantitative, log-linear cognitive diagnosis modeling	Ability classification

Table A1 (Continued). Complete TCM classification matrix

No	Study	Country	Theory (T)	Context (C)	Method (M)	Research focus
26	Paulsen and Valdivia (2021)	USA	CDM	General/methodological	Quantitative, CDM	Ability classification
27	Li et al. (2021)	China	CDM	Reading, primary	Quantitative, CDMs	Ability classification
28	Hu et al. (2021)	China	CDM	General/methodological, primary	Quantitative, DINA model	Ability classification
29	Tian et al. (2020)	China	CDA	General/methodological	Quantitative, Monte Carlo simulation	Ability classification
30	Lin et al. (2020)	USA	CDM	General/methodological	Quantitative, latent growth curve modeling	Ability classification
31	Mirzaei et al. (2020)	Iran	CDM	General/methodological	Quantitative, G-DINA	Ability classification
32	Chin et al. (2020)	Malaysia	Theory of learning from error proposed	General/methodological, primary	Mixed-methods, thematic analysis	Ability classification
33	Wu et al. (2020)	China	CDM	Mathematics	Quantitative, CDMs	Ability classification
34	Nájera et al. (2019)	Spain	CDM	General/methodological	Quantitative, discrimination index	Ability classification
35	Ma and de la Torre (2019)	USA	CDM	General/methodological	Quantitative, G-DINA	Ability classification
36	Wang et al. (2019)	China	CDM	General/methodological	Quantitative, expected classification accuracy index	Ability classification
37	da le Torre et al. (2018)	Brazil	Nan	General/methodological	Quantitative, Bayesian estimation via the no-U-turn sampler algorithm	Q-matrix validation
38	Hung and Huang (2018)	Taiwan	CDM	General/methodological	Quantitative, Bayesian estimation	Ability classification
39	Kang et al. (2018)	China	CDA	Language	Quantitative, RMSEA	Ability classification
40	Shih et al. (2018)	Taiwan	CDM	General/methodological	Quantitative, higher-order DINA model	Ability classification
41	Wang and Qiu (2018)	USA	CDM	General/methodological	Mixed-methods, Bayesian estimation with MCMC methods	Ability classification
42	Ravand and Robitzsch (2018)	Iran	CDM	Reading	Quantitative, model fit indices	Ability classification
43	Wu, 2018)	Taiwan	nan	Mathematics	Quantitative, one-way ANCOVA	Ability classification
44	Aryadoust (2018)	Singapore	CDM	Listening	Quantitative, CDA	Ability classification
45	Tu et al. (2018)	China	CDM	General/methodological	Quantitative, MLE	Ability classification
46	Yamaguchi and Okada (2018)	Hong Kong	CDM	Mathematics	Quantitative, CDMs	Ability classification
47	Köhn and Chiu (2018)	USA	Lattice theory is used …	General/methodological	Quantitative, Boolean operations and lattice theory	Ability classification
48	Park et al. (2017)	USA	CDM	General/methodological	Quantitative, latent GOLD 5.0	Ability classification
49	Ranjbaran and Alavi (2017)	Iran	CDM	Reading	Quantitative, RUM analysis	Ability classification
50	Lei and Li (2016)	USA	CDM	General/methodological	Quantitative, G-DINA	Ability classification
51	Liu et al. (2016)	USA	The hierarchical design ...	General/methodological	Quantitative, hierarchical diagnostic classification model	Ability classification
52	Kuo et al. (2016)	Taiwan	CDM	Mathematics, primary	Quantitative, expectation maximization algorithm	Ability classification

Table A1 (Continued). Complete TCM classification matrix

No	Study	Country	Theory (T)	Context (C)	Method (M)	Research focus
53	Skaggs et al. (2016)	USA	CDM	Mathematics, secondary	Quantitative, CDM	Ability classification
54	Cui et al. (2015)	Canada	Neural networks	General/methodological	Quantitative, artificial neural networks	Ability classification
55	Leighton and Chu (2015)	Canada	-	General/methodological	Quantitative, none specified	Ability classification
56	Madison and Bradshaw (2014)	USA	CDM	General/methodological	Quantitative, log-linear cognitive diagnosis model	Ability classification