Journal of Humanities and Social Sciences Studies
ISSN: 2663-7197
DOI: 10.32996/jhsss
Journal Homepage: www.al-kindipublisher.com/index.php/jhsss

JHSSS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

A Review of Cognitive Diagnosis Assessment in Language from 2000 to 2021: A Visualized Analysis with CiteSpace II

Huihui Sun[1] ✉ and Yuying Kang[2]

[12]Department of English Language and Culture, Jeonju University, Jeonju 55069, Korea
[1]Department of Foreign languages, Lyuliang University, Lyuliang 033000, China
[2]Department of Business English, Foreign languages school of Shanxi Datong University, Datong, 037009, China

Corresponding Author: Huihui Sun, **E-mail**: 372741616@qq.com

| ABSTRACT

As an assessment approach, Cognitive Diagnostic Assessment (hereafter CDA) provides fine-grained evaluations of examinees based on their test performance for stakeholders. It has received much attention since it was applied in language testing. Hence, it is crucial to keep abreast of emerging trends and the intellectual base of CDA in language assessment to offer guidance for future testers. In order to have a whole picture of CDA in language, a synthesized network is depicted in this paper based on 1614 original research and review papers, which are obtained from a refined topic search on "cognitive diagnosi*" from 2000 and 2021. CiteSpace is utilized to simplify the analysis of the research hotspots, emerging trends, and intellectual structure. The paper starts with a brief description of the involved time, region, and discipline, aiming to picture a whole image of CDA. Then research topics and the intellectual structure of CDA are emphasized. The findings show that most existing research focuses on theoretical discussions. Few practical applications highlight writing diagnoses and retrofitting reading assessments. A recent emerging trend is utilizing CDA to implement longitudinal supervision through computer-based assessments and intelligent tutoring systems.

| KEYWORDS

CiteSpace II, Cognitive Diagnostic Assessment, Language Assessment

| ARTICLE INFORMATION

ACCEPTED: 23 August 2022 **PUBLISHED:** 25 August 2022 **DOI:** 10.32996/jhsss.2022.4.3.17

1. Introduction

Cognitive Diagnostic Assessment (CDA) could provide current skill mastery probabilities of examinees for test score users. Compared with the general ability level information provided by the proficiency and achievement assessments, CDA has a significant advantage in that it increases the information value and interpretation power of examinees' scores by providing their strengths and weaknesses. In this way, testing is not just used for evaluating and ranking test-takers in a group but for guiding teaching and learning. Teachers and students could take appropriate actions to remedy learners' weaknesses based on the fine-grained information provided by the CDA. Because CDA plays a vital role in education, numerous studies of CDA have been conducted in various fields. This paper aims to detect the usage situation of CDA in the language area.

It is not easy to depict a whole picture of a specific area, especially when CDA is an interdisciplinary field. Reviews of CDA in language assessment were absent, and few existing articles introduce the general usage of CDA in language testing. For example, an overview by Lee and Sawaki (2009) focuses on procedures of CDA, differences between psychometric models, and challenges and avenues for the future language CDA. These exploratory studies did not indicate development trends of CDAs and could not offer sufficient references for language testers. A comprehensive survey based on a massive amount of data is needed. Therefore, this paper aims to present the research focus, emerging trends, and intellectual structure of CDA in language with CiteSpace II.

CiteSpace II is a tool for knowledge mapping. It allows the visualization of trends in a particular knowledge area by integrating vast amounts of abstract data into interactive visual representations. Therefore, it has been used in various filed for various purposes. Chen et al. (2012) proved with CiteSpace II that the studies of stem cells played a significant role in regenerative medicine. In the same year, the researcher who made an outstanding contribution to the development of stem cells received the Nobel Prize.

The paper aims to analyze CDA in language assessment with CiteSpace II quantitatively. The article describes the whole picture, emerging trends, and intellectual base of CDA in language assessment. Specifically, the paper presents the region, time, and countries involved in language CDAs, then research hotspots and the newest emerging trends and intellectual structures are further pointed out. The paper ends up with limitations and suggestions for future research. The following questions guide the paper：

(1) What are the research focuses and emerging trends of language CDA?
(2) What are the intellectual structures of language CDA?

2. CiteSpace II

CiteSpace II is a Java application developed by Drexel University, USA. The main goal of this instrument is to analyze and visualize the emerging trends in a certain field. This tool also offers several options to interpret historical patterns of a knowledge domain and provides the intellectual base by detecting the primary citations and clustering them.

Like other bibliometric networks, the main characteristic of CiteSpace networks is that they are composed of nodes and edges. Depending on outputted graphics, nodes represent different contents, such as keywords, authors, countries, and publications. Edges indicate the relations of nodes and the strengths of their connections.

CiteSpace II depicts the literature with a synthetic network comprising a series of individual networks. Each separate network is established by articles published in a time interval, known as the time slice. For this review, the time slice is 1, which means a network is formed each year. Then these individual networks are connected through the same nodes involved in them. The threshold values represent how many nodes are included in each separate network. The threshold of this paper is Top 50, which means the top 50 high-frequency nodes each year are taken to form networks. Then the whole network is connected by the same nodes in individual networks.

3. Methodology

This paper used CiteSpace II to generate and analyze various mappings of CDA, namely, the time growth diagram, national cooperation map, subject category map, keyword co-occurrence network, and citation clusters. They jointly depicted an overview picture of CDA. The original data was extracted from the Web of Science core collection. Because the body of the relevant literature grows every year, in this article, the literature was reviewed as of October 2021. An exact topic search for "cognitive* diagnos*" in titles, abstracts, or indexing terms resulted in 805 records from 2000 to 2021. After filtering out less representative record types such as proceedings papers and notes, the dataset was reduced to 665 original research and review articles. The remaining 665 records were manually checked one by one through titles and abstracts to improve the accuracy of the search. Unrelated records were excluded, such as CDA in anxiety disorders and mathematics problems. Finally, 438 papers on theoretical research and practical applications of CDA in language assessments were left.

Given that CiteSpace forms a visual map based on the data cluster, the data should reach a certain size (Chen, 2016). The dataset was expanded by citation indexing. The article, which cites at least one of the 438 records, was included in the expanded dataset based on the assumption that citing an article about cognitive diagnosis makes the citing article relevant to the topic. The citation expansion resulted in 1,176 records. Self-citation was removed. The time range of the expanded dataset remains 2000-2021. Thus, the analysis focuses on language CDA over the last twenty years. The 1614 article dataset was used in the subsequent analysis.

4. Results and Discussion

4.1 The Distribution of Cognitive Diagnosis

Based on 1614 retrieved literature, this section examines the distribution of related literature from time, region, and discipline to picture a whole image of language CDA. They are shown respectively in Figure 1, Figure 2, and Figure 3.

Figure 1. *Time Diagram of CDA*

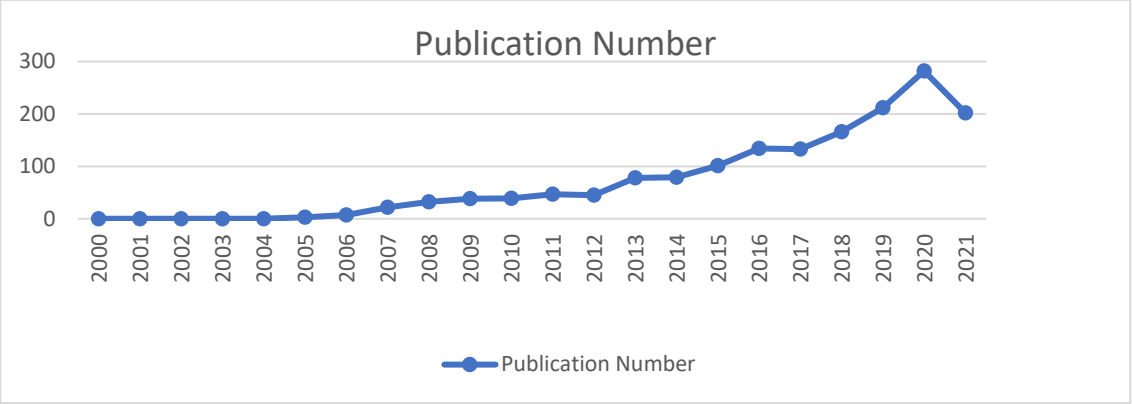


Figure 1 shows that during the observation period (2000-2021), the quantity of relevant literature increased steadily, which means the popularity of CDA was on the rise. Scholars have realized the importance of offering stakeholders specific feedback on examinees' test scores.

Figure 2. *Countries in CDA between 2000-2021*

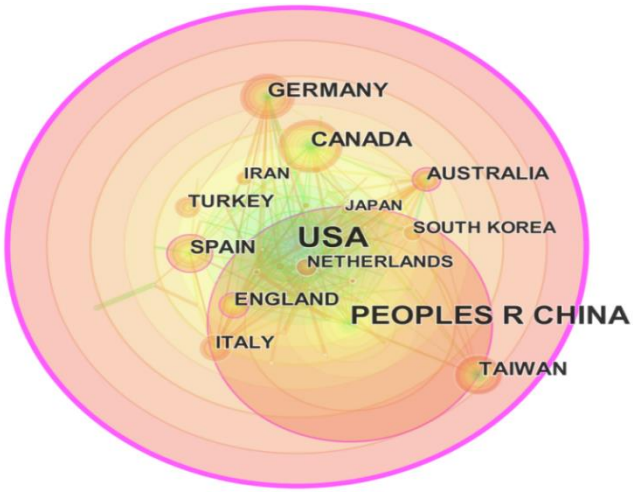
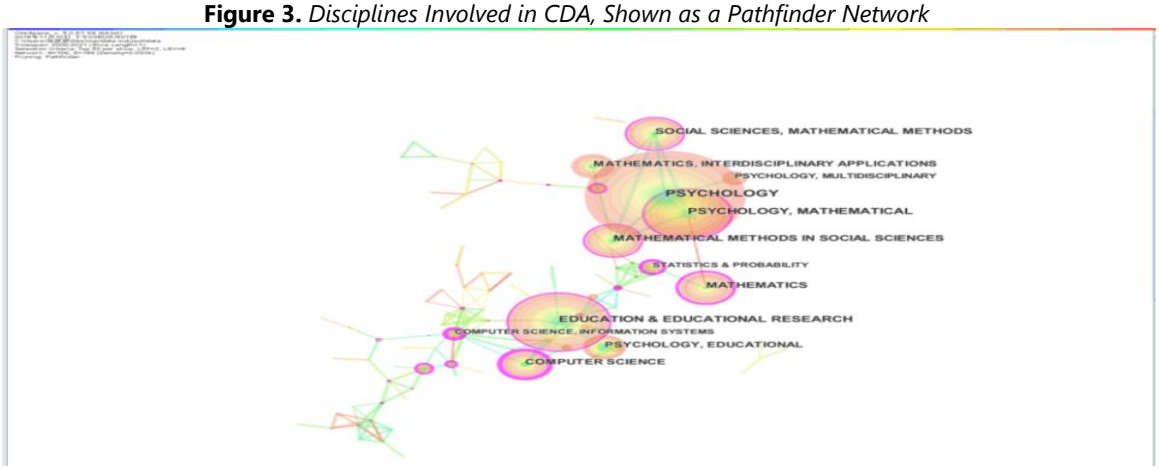


Figure 2 is the science mapping of research countries. The countries are labeled as nodes. The sizes of nodes agree with the number of publications in corresponding countries. The larger the node is, the more publications the country has. A purple ring around a node represents the betweenness centrality of the country. Betweenness centrality refers to how closely the country cooperates with other countries. The thicker the purple circle is, the closer the country is to other countries. Rings in red indicate that publication bursts are detected in these countries.

As shown in Figure 2, many countries have conducted studies on CDA. Among them, *America* had the most publications, followed by *China*. Although *Canada, German, Spain, and South Korea* were much smaller, they had sufficient research to show in the Figure. Besides, the red color of the whole Figure 2 indicates that language CDA is a relatively new area. It is consistent with the documentation. When CDA was put forward, it was first used in the medical area to treat diseases, such as anxiety disorder. It was applied in language assessment until 1997 by Buck et al.

It is worth noting that the node of *China* is almost in red, which shows studies of CDA in language testing burst in China in recent years. Moreover, the large size of the *China* Node shows that since CDA was brought in, many types of research have been conducted by Chinese scholars. The thick purple circle around the *American* node shows that America, as one of the origin countries, has a close relationship with others. The purple rings around England, Spain, and Australia mean that these three countries also play the bridge role in the CDA network.

Figure 3. *Disciplines Involved in CDA, Shown as a Pathfinder Network*

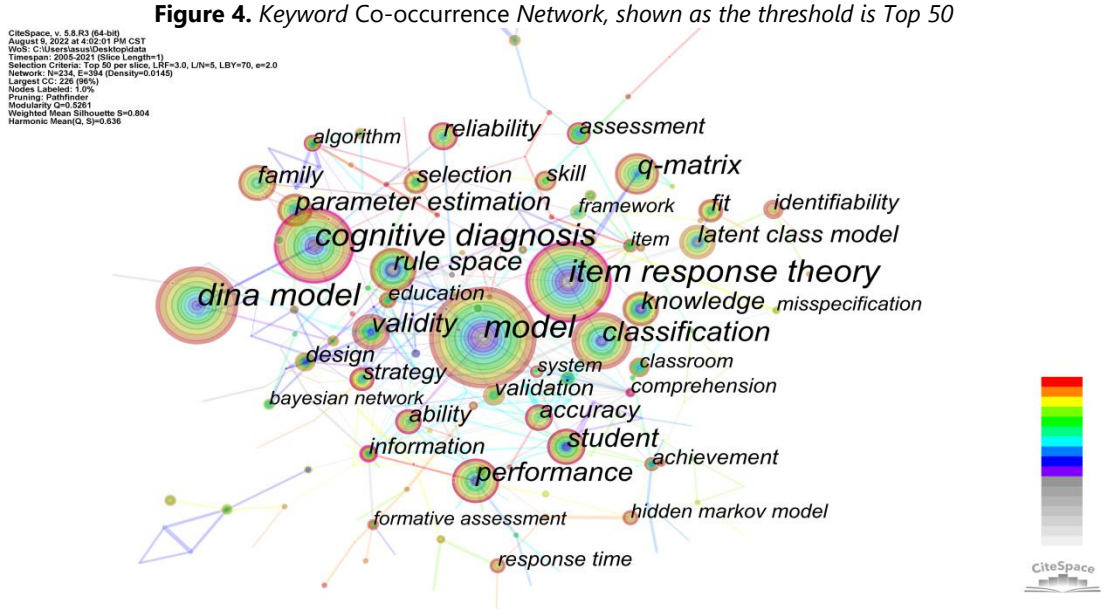


Which disciplines are involved in CDA? Each article acquired from the Web of Science is assigned one or more subject categories. Figure 3 shows a network of subject categories after being simplified by Pathfinder, with which the most prominent connections could remain. The most common type was *Psychology*. *Education & Education Research* was the second, proving CDA has been extensively applied in education. However, a separate classification of language education was missing indicating that CDA has not been used in language instruction widely. Extrusive purple rings around *Computer Science, Mathematics, and Statistics & Probability* indicate that CDA is a highly interdisciplinary field. Because CDAs require high computer skills, it is reasonable to infer that the technical difficulty is one of the limitations of the applications of CDA in language testing.

4.2 Research Topics in Recent 20 Years

CiteSpace II offers the keyword co-occurrence mapping to capture the research topics. Because keywords are the representatives of the publications, a keyword co-occurrence network could signify a specific area's pivotal points and research fronts. In order to keep the completeness of the research topics, the following picture takes the TOP 50 as the threshold, which means that the top 50 high-frequency keywords in each year are taken to form the network. Pathfinder's link reduction algorithm was chosen to maintain the readability of the map, and the merged network was selected to keep the completeness of the information. There are two link reduction algorithms in CiteSpace II: Minimum spanning trees (MSTs) and Pathfinder networks (PENETs). Compared with MST, PENET maintains the cohesiveness of the pivotal paths and makes mapping more predictable and interpretable (Chen & Morris, 2003). For this purpose, the study chose PENET as the reduction method. Figure 4 shows the keyword co-occurrence mapping considering all these things. In the following explanations, the nodes with big size or in red are emphasized because they represent the significant research points and the emerging trends, respectively.

Figure 4. *Keyword* Co-occurrence *Network, shown as the threshold is Top 50*



Through organizing the keywords in Figure 4 and further reading original articles represented by the keywords, the study found that the development of CDA focuses on the following directions: 1) developing new psychometric models, 2) finding new ways to confirm the accuracy of Q-matrices, and 3) practical applications of CDA in language. The findings proved that most existing studies were on the theoretical construction of CDA. Only a few academic achievements were used in the practice of language assessment. Detailed analyses for each direction are presented below to guide future language testers.

Nodes *latent class model, DINA model, diagnostic classification model*, and *cognitive diagnosis model* have a large size in the network, which indicates that the development of cognitive diagnostic models (hereafter CDMs) has been valued since CDA was proposed. According to the original publications, more than 100 models are available as of 2020. These models are classified into non-compensatory, compensatory, and general models based on their model assumptions. Take English reading as an example to explain the different model assumptions. If three reading skills are required to finish a reading test successfully, non-compensatory models ask examinees to grasp all three reading abilities. However, compensatory models assume skills are complementary. A highly mastered skill can make up for an unmastered skill. However, it is often difficult to determine whether skills are complementary because of the complex nature of a specific field. If the chosen model is unsuitable, it leads to severe consequences. Along this line, when most CDMs are compensatory or non-compensatory, de la Torre developed a general model in 2011, namely, the Generalized Deterministic Inputs, Noisy and Gate Model (G-DINA). As a general model, the G-DINA model has broader applicability because of its more relaxed model assumptions. G-DINA can accept non-compensation and compensation relationships between attributes at the same time. Besides, Templin and Bradshaw (2013) introduced the Hierarchical Diagnostic Classification Model (HDCM), which advanced psychometric models in their capacity to evaluate the presence of attribute hierarchies objectively.

Nodes *Accuracy, Validity, Reliability, and Q-matrix* show that except for appropriate CDMs, accurate Q-matrix is also a critical factor in ensuring the accuracy of diagnostic results. If misspecifications of the Q-matrix are left unchecked, it results in wrong interpretations of diagnostic information. Various means and indices are purposed to build and examine Q-matrix, such as substantive information about the items, expert knowledge about the domain, and verbal protocols from students. For further explanation, some significant publications are listed in the following. De la Torre (2008) proposed a statistical method. He proved with simulated and empirical data that the sequential EM-based method could be used in conjunction with other ways to validate the appropriateness of the Q-matrix. Afterward, Chen et al. (2013) proposed relative and absolute fit for evaluating the misfits of CDM and Q-matrix, or both misfits. They investigated the sensitivity of various fit indexes under different misfit settings. De la Torre and Chiu (2016) proposed a general discrimination index, namely."s^2" to avert the subjective construction of the Q-matrix. New calibrations were implemented as well to improve the accuracy of the Q-matrix. Chen et al. (2015) suggested the regularized maximum as an estimation procedure for constructing Q-matrix based on the DINA and the DINO models.

Nodes *comprehension, student,* and *instruction* signify that CDA has been used in writing and reading comprehension assessments on a small scale. Related literature of these keywords shows that CDA has been used in reading comprehension of the second and foreign language, whereas most publications were retrofitting CDAs. There are two methods of CDAs. The first is developing a new diagnostic testing, such as DIALANG. The other is extracting diagnosis information from existing non-diagnostic testing. Most existing CDAs took the second method and retrieved diagnostic information from large-scale international English ability assessments, such as TOEFL and IELTS (Clark & Endres, 2021; Gao & Rogers, 2011; Jang, 2009). Only Ranjbaran and Alavi (2017) explored how to develop a reading comprehension test for diagnostic purposes. Besides, the eye-catching purple circle around the *Comprehension* Node indicated that reading comprehension had played a connection role in the applications of CDA in language.

4.3 Intellectual Structure of CDA

All cited references form the intellectual structure of a research field. CiteSpace further classifies these references into several clusters. Articles with similar content are grouped into one category. In the following, the first eight clusters are listed first, then five remarkable clusters are explained in detail.

Table 1: *Clusters of Cited References*

Cluster-ID	Size	Silhouette	Year Ave.	Label
0	66	0.814	2012	Diagnostic Testing
1	34	0.813	2005	Component
2	30	0.848	2008	Computer Based Testing
3	28	0.974	2006	Cognitive Load
4	28	0.908	2003	Skills Diagnosis
5	25	0.906	2014	Longitudinal Cognitive Diagnostic Assessment
6	17	1	2000	Cognitive Diagnosis
8	13	0.916	2007	Language Assessment

Eight primary clusters are listed in Table 1. The quality of a cluster is reflected by its silhouette score, which should be close to 1. As shown in Table 1, all the clusters are highly homogeneous with a high silhouette value. The average year of a cluster indicates its recentness. The later the time is, the newer the research is. For example, longitudinal cognitive diagnostic assessment in cluster #5 was the latest research topic in the applications of CDA in language testing because its average formation time was the newest. Besides, each cluster is given a label automatically by extracting noun phrases from the titles of citing articles. Three algorithms (LSI, LLR, MI) are available in this step. Among them, LLR is recommended by the primary developer of CiteSpace because it can maintain the accuracy of extraction to the greatest extent. Therefore, labels chosen by LLR are used in the subsequent discussions.

In order to offer the most direct references for future language testers, five clusters (#0, #1, #2, #5, #8) are explained in detail. Five representative citing articles and cited references for each cluster were selected and highlighted. Besides, this study finds that Clusters #3 and #4 appear to almost be formed by the publications of Kalyuga and de la Torre, respectively. Readers interested in cognition load and skills diagnosis can further consult their articles.

Table 2: *Cluster #0 Diagnostic Testing*

Cited Reference		Citing Article	
Cite	Author (Year)	Coverage %	Author (Year) Journal, Volume
178	De la Torre (2011)	18	Wen et al. (2020)
104	Henson et al. (2009)	18	Ravand & Baghaei (2020)
80	De la Torre & Chiu (2016)	17	Ma & De la Torre (2020b)
77	Chen et al. (2015)	15	Yu & Cheng (2019)
66	Decarlo (2011)	11	Ma & De la Torre (2020a)

The core cited references of Cluster #0 are the significant milestones of the development of diagnostic testing. Notably, De la Torre's (2011) *Generalized DINA Model Framework* proposed the G-DINA model to generalize the DINA model and verify that specific CDM formulations and G-DINA formulations could be interchanged when appropriate constraints are applied. Henson et al. (2009) defined a family of CDMs using log-linear models with latent variables. They further discussed the relationship between many common models. De la Torre and Chiu (2016) and Chen et al. (2015) refined the checking of the Q-matrix.

The citing articles with the high coverage of the intellectual structure in this cluster further developed CDMs based on previous models. For example, Wen et al. (2020) developed longitudinal CDMs to evaluate longitudinal growth in skills mastery. Ravand and Baghaei (2020) concluded the latest developments of CDMs, and proposed suggestions to make CDMs work smoothly and quickly in educational systems. Ma and De la Torre (2020a) evolved the validation of the Q-matrix from dichotomous responses to graded response data.

Table 3: *Cluster #1 Component*

Cited reference		Citing article	
Cite	Author (Year)	Coverage %	Author (Year) Journal, Volume
36	Leighton & Gierl (2007)	20	Almond (2010)
34	Jang (2009)	15	Sinharay et al. (2010)

33	Leighton et al. (2004)	15	Robusto et al. (2010)
22	Roussos et al. (2007)	12	Leighton et al. (2010)
16	Almond et al. (2007)	12	Huff et al. (2010)

Cluster #1 is formed relatively earlier than other clusters. The prominent members of this cluster introduced CDA for education. For example, the book edited by Leighton and Gierl (2007) includes 12 remarkable articles about the basis of CDA, test design and analysis principles, and psychometric procedures and applications. Roussos et al. (2007) presented an overarching framework of the latent class approach based on the broad review, and they further pointed out the development status and deficiencies of CDA.

The published time of primary citing articles in this cluster is also earlier. They are preliminary studies on the use of cognitive diagnostics in education. Sinharay et al. (2010) studied existing temptations, pitfalls, and some solutions to reporting diagnostic scores in educational testing.

Table 4: *Cluster #2 Computer-Based Testing*

Cited reference		Citing article	
Cite	Author (Year)	Coverage %	Author (Year)
48	Cheng (2009)	23	Barrada (2012)
38	Dibello et al. (2007)	23	Chang (2012)
36	Reckase (2009)	13	Lawrence (2014)
30	Henson & Douglas (2005)	13	Wang et al. (2012)
22	Embretson & Yang (2013)	10	Wang et al. (2011)

Cluster #2 was also formed earlier, whose principal members were used as references for item stratification in CD-CAT. Cheng (2009) first showcased the application of the optimal sequential selection methodology in item selection of CD-CAT. As of December 2021, this article had 128 citations. In the fourth highly cited article, *Test Construction for Cognitive Diagnosis,* Henson and Douglas (2005) proposed a general CDM index (CDI.) to improve the discrimination of items among examinees by calibrating the accuracy of item selection. This index offered a guideline for constructing a good test with CDMs, effectively avoiding random test construction. This paper was cited by Wang et al. (2011, 2012) to create an item selection approach in the CD-CAT.

Table 5: *Cluster #5 Longitudinal Cognitive Diagnostic Assessment*

Cited reference		Citing article	
Cite	Author (Year)	Coverage%	Author (Year)
47	Li et al. (2016)	36	Zhan (2020a)
40	de la Torre & Minchen (2014)	32	Zhan (2020b)
39	George et al. (2016)	28	Lin et al. (2020)
33	Wang et al. (2018)	28	Pan et al. (2020)
32	Kaya & Leite (2017)	24	Wen et al. (2020)

Cluster #5 is the newly formed cluster that hides the research frontier of the CDA. The publications in this cluster indicate that longitudinal cognitive diagnostic assessment is the future direction of CDA. The article by Li et al. (2016) demonstrated the possibility of using the DINA model in the dynamic supervision of the changes in cognitive skills over time. Wang et al. (2018) proposed a new framework that integrates one of the cognitive diagnosis models with the hidden Markov model to trace the students' skill transition in the learning environment. Kaya and Leite (2017) presented longitudinal cognitive diagnosis modeling, which can be used to monitor the attribute stability of individuals through repeated measurements.

The five major citing articles were all published in 2020. Moreover, all of them take longitudinal learning diagnosis as the topic. The future direction of CDA is the longitudinal diagnosis from this newly formed intellectual structure.

Table 6: *Cluster #8 Language Assessment*

Cited reference		Citing article	
Cite	Author (Year)	Coverage%	Author (Year)
14	Lee & Sawaki (2009)	30	Lee & Sawaki (2010)
10	Hattie & Timperly (2007)	30	Alderson (2010)
9	Alderson et al. (2015)	23	Sawaki et al. (2009)
8	Sawaki et al. (2009)	23	Jang (2009)
8	Alderson (2005)	15	Harding et al. (2015)

Cluster #8 is a cluster on the theoretical basis and practical applications of CDA in language. The second highly cited work, *The Power of Feedback,* written by Hattie and Timperley in 2007, had 3975 citations on the Web of Science until December 2021. It emphasizes that feedback is an effective method of identifying gaps between task, process, and self-regulatory. Although it does not refer to CDA, it affirms the value of remedial strategies. This is exactly what CDA emphasizes. When scholars extracted diagnosis information from standardized proficiency tests, Alderson et al. (2015) began to notice the procedures of diagnostic language assessment. They proposed five principles of diagnostic language assessment. In 2009, Sawaki et al. took the Test of English as a Foreign Language ™ Internet-based Test as the testing instrument and focused on the core language skills assessed in the reading and listening sections of the test. The citing article by Harding et al. (2015) developed a tentative framework for a diagnosis theory in second or foreign language assessment when most existing diagnostic tests were retrofitting.

5. Conclusion

This paper has outlined the evolutionary trajectory of CDA in English over the last twenty years. Initially, this paper depicts a whole picture of the applications of CDA in English. The findings show that scholars have paid considerable attention to CDA, and various countries and disciplines are involved in this field. The second aim of the study was to identify the research topics. Three research directions are detected: developing new models, building and proofreading Q-matrix, and practical applications in English. The other aim was to discuss the intellectual structures of the CDA. This paper retrieved eight clusters and explained five of them.

The above results reveal some problems with CDA. First, applications of CDA in language are absent, and the existing studies are retrofitting. Moreover, retrofitting studies have inner limitations. The distractors in actual cognitive diagnostic tests are designed for a particular type of error. The choice of a specific distractor reflects a problem with understanding a concept or using the rule. However, the retrofitting CDAs cannot achieve this requirement. The other shortcoming is existing CDAs lack follow-up teaching. Besides, existing CDAs focus on English reading and writing. CDAs on English listening, speaking, and translation are absent.

6. Limitations and Suggestions

Although the dataset was expanded by citation indexing, this paper only focused on the references in the WoS dataset. Data in other databases were not analyzed. CiteSpace II can also directly import data from the arXiv database and provides format converters for data derived from CNKI, CSSCI, Derwent, NSF, SCOPUS, SDSS, and Project DX. Future researchers can expand data retrieval methods to accurately and comprehensively retrieve all literature on research topics. In addition, this paper did not display development trends on CDA over time. Future researchers can pay attention to it with Timeline and Timezone views of CiteSpace II.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* A&C Black.

[2] Alderson, J. C. (2010). Cognitive diagnosis and Q-Matrices in language assessment: A commentary. *Language Assessment Quarterly, 7*(1), 96-103. https://doi.org/10.1080/15434300903426748

[3] Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics, 36*(2), 236-260. https://doi.org/10.1093/applin/amt046

[4] Almond, R. G. (2010). Using evidence-centered design to think about assessments. In V. J. Shute & B. J. Becker (Eds.), *Innovative Assessment for the 21st Century: Supporting Educational Needs* (pp. 75-100). Springer US. http://doi.org/10.1007/978-1-4419-6530-1_6

[5] Almond, R. G., Dibello, L. V., & Moulder, B. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement, 44*(4), 341-359. http://doi.org/10.1111/j.1745-3984.2007.00043.x

[6] Barrada, J. R. (2012). Computerized adaptive testing a general perspective. *Anales de Psicologia, 28*(1), 289-302.

[7] Buck, G., Tatsuoka, K., & Kostin, I. (1997). The subskills of reading: Rule-space analysis of a multiple-choice test of second language reading comprehension. *Language Learning*, *47*(3), 423-466. http://doi.org/10.1111/0023-8333.00016

[8] Chang, H. H. (2012). Making computerized adaptive testing diagnostic tools for s chool. In R. W. Lissitz & H. Jiao (Eds.), *Computers and Their Impact on State Assessments: Recent History and Predictions for the Future* (pp. 195-226). Information Age Publishing.

[9] Chen, C. (2016). *CiteSpace: A practical guide for mapping scientific literature*. Nova Science Publishers.

[10] Chen, J. S., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement, 50*(2), 123-140. https://doi.org/10.1111/j.1745-3984.2012.00185.x

[11] Chen, C. M., Hu, Z. G., Liu, S. B., & Tseng, H. (2012). Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace. *Expert Opinion on Biological Therapy, 12*(5), 593-608. https://doi.org/10.1517/14712598.2012.674507

[12] Chen, Y. X., Liu, J. C., Xu, G. J., & Ying, Z. L. (2015). Statistical analysis of Q- matrix based diagnostic classification models. *Journal of the American Statistical Association, 110*(510), 850-866. http://doi.org/10.1080/01621459.2014.934827

[13] Chen, C.M., & Morris, S. (2003). Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In T. Munzner & S. North (Eds.), *IEEE Symposium on Information Visualization 2003* (pp.67-74). IEEE. http://doi.org/ 10.1109/INFVIS.2003.1249010

[14] Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika, 74*(4), 619-632. http://doi.org/10.1007/s11336-009-9123-2

[15] Clark, T., & Endres, H. (2021). Computer-based diagnostic assessment of high school students' grammar skills with automated feedback–an international trial. *Assessment in Education: Principles, Policy & Practice, 28*(5-6), 602-632. https://doi.org/10.1080/0969594X.2021.1970513

[16] DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement, 35*(1), 8-26. https://doi.org/10.1177/0146621610377081

[17] De la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*(4), 343-362. https://doi.org/10.1111/j.1745-3984.2008.00069.x

[18] De la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179-199. http://doi.org/10.1007/S11336-011-9207-7

[19] De la Torre, J., & Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika, 81*(2), 253-273. http://doi.org/10.1007/s11336-015-9467-8

[20] De la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicologia Educativa, 20*(2), 89-97. http://doi.org/10.1016/j.pse.2014.11.001

[21] Dibello, L. V., Roussos, L. A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of Statistics, 26,* 979-1030. http://doi.org/10.1016/S0169-7161(06)26031-0

[22] Embretson, S. E., & Yang, X. D. (2013). A multicomponent latent trait model for diagnosis. *Psychometrika, 78*(1), 14-36. http://doi.org/10.1007/s11336-012-9296-y

[23] Gao, L., & Rogers, W. T. (2011). Use of tree-based regression in the analyses of L2 reading test items. *Language Testing, 28*(1), 77-104. https://doi.org/10.1177/0265532210364380

[24] George, A. C., Robitzsch, A. & Kiefer, T. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software, 74*(2), 1-24. http://doi.org/10.18637/jss.v074.i02

[25] Harding, L. Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing, 32*(3), 317-336. http://doi.org/10.1177/0265532214564505

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

[27] Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement, 29*(4), 262-277. http://doi.org/10.1177/0146621604272623

[28] Herson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191-210. http://doi.org/10.1007/S11336-008-9089-5

[29] Huff, K., Steinberg, L. & Matts, T. (2010). The promises and challenges of implementing evidence-centered design in large-scale assessment. *Applied Measurement in Education, 23*(4), 310-324. http://doi.org/10.1080/08957347.2010.510956

[30] Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing, 26*(1), 31-73. http://doi.org/10.1177/0265532208097336

[31] Kaya, Y., & Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educational and Psychological Measurement, 77*(3), 369-388. http://doi.org/10.1177/0013164416659314

[32] Lee, Y. W., & Sawaki, Y. (2009). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly, 6*(3), 172-189. http://doi.org/10.1080/15434300902985108

[33] Lee, Y. W., & Sawaki, Y. (2010). Cognitive diagnosis and Q-matrices in language assessment: The authors respond. *Language Assessment Quarterly, 7*(1), 108-112. http://doi.org/10.1080/15434300903559076

[34] Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.

[35] Leighton, J. P., Gierl, M. J., Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement, 41*(3), 205-237. http://doi.org/10.1111/j.1745-3984.2004.tb01163.x

[36] Leighton, J. P., Gokiert, R. J., & Cor, M. K. (2010). Teacher belief in the cognitive diagnostic information of the classroom versus large-scale tests. *Assessment in Educational-Principles Policy & Practice, 17*(1), 7-21. http://doi.org/10.1080/09695940903565362

[37] Li, F. M., Cohen, A., Bottge, B., & Templin, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement, 76*(2), 181-204. http://doi.org/10.1177/0013164415588946

[38] 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,* 2205. http://doi.org/10.3389/fpsyg.2020.02205
[39] Ma, W., & de la Torre, J. (2020a). 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
[40] Ma, W., & de la Torre, J. (2020b). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, *93*(14), 1-26. https://doi.org/10.18637/jss.v093.i14
[41] Pan, Q. Q., Qin, L., & Kingston, N. (2020). Growth modeling in a diagnostic classification model (DCM) framework-A multivariate longitudinal diagnostic classification model. *Frontiers in Psychology,11,* 1714. http://doi.org/10.3389/fpsyg.2020.01714
[42] Ravand, H., & Baghaei, P. (2020). Diagnostic classification models: Recent developments, practical issues, and prospects. *International Journal of Testing, 20*(1), 24-56. https://doi.org/10.1080/15305058.2019.1588278
[43] 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. http://doi.org/10.1016/j.stueduc.2017.10.007
[44] Robusto, E., Stefanutti, L. & Anselmi, P. (2010). The gain-loss model: A probabilistic skill multimap model for assessing the learning process. *Journal of Educational Measurement, 47*(3), 373-394. http://doi.org/10.1111/j.1745-3984.2010.00119.x
[45] Roussos, L. A., Templin, J. L., Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement, 44*(4), 293-311. http://doi.org/10.1111/j.1745-3984.2007.00040.x
[46] Reckase, M. D. (2009). Multidimensional item response theory models. In *Multidimensional item response theory model* (pp. 79-112). Springer, New York.
[47] Sawaki, Y., Kim, H. J., & Gentile, C. (2009). Q-Matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Language Assessment Quarterly, 6*(3), 190-209. http://doi.org/10.1080/15434300902801917
[48] Sinharay, S., Puhan, G., & Haberman, S. J. (2010). Reporting diagnostic scores in educational testing: Temptations, pitfalls, and some solutions. *Multivariate Behavioral Research, 45*(3), 553-573. http://doi.org/10.1080/00273171.2010.483382
[49] Templin, J., & Bradshaw, L. (2013). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika, 79*(2), 317-339. http://doi.org/10.1007/S11336-013-9362-0
[50] Wang, C., Chang, H. H., & Douglas, J. (2012). Combining CAT with cognitive diagnosis: A weighted item selection approach. *Behavior Research Methods, 44*(1), 95-109. http://doi.org/10.3758/s13428-011-0143-3
[51] Wang, C., Chang, H. H., & Huebner, A. (2011). Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement, 48*(3), 255-273. http://doi.org/10.1111/j.1745-3984.2011.00145.x
[52] Wang, S. Y., Yang, Y., Culpepper, S. A., Douglas, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order, hidden Markov model with covariates. *Journal of Educational and Behavioral Statistics, 43*(1), 57-87. http://doi.org/10.3102/1076998617719727
[53] Wen, H. B., Liu, Y. P., & Zhao, N. N. (2020). Longitudinal cognitive diagnostic assessment based on the HMM/ANN model. *Frontiers in Psychology, 11,* 2145. http://doi.org/10.3389/fpsyg.2020.02145
[54] Yu, X., & Cheng, Y. (2020). Data-driven Q-matrix validation using a residual-based statistic in cognitive diagnostic assessment. *British Journal of Mathematical and Statistical Psychology, 73*(S1), 145-179. https://doi.org/10.1111/bmsp.12191
[55] Zhan, P. D. (2020a). A Markov estimation strategy for longitudinal learning diagnosis: Providing timely diagnostic feedback. *Educational and Psychological Measurement, 80*(6), 1145-1167. http://doi.org/10.1177/0013164420912318
[56] Zhan, P. D. (2020b). Longitudinal learning diagnosis: Minireview and future research directions. *Frontiers in Psychology, 11,* 1185. http://doi.org/10.3389/fpsyg.2020.01185