BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

A Framework for Constructing Cognitive Diagnostic Reports on English Reading for Chinese University Students

Zhe Zhang [*] **, *Mohd Isa Hamzah*** **, *Khairul Azhar Jamaludin***

Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia

ABSTRACT

In the context of higher education in China, improving students' English language proficiency has become a critical task, as English plays an increasingly important role in academic development and career success. However, traditional summative assessment methods, such as the College English Test Band 4 and Band 6 (CET-4 and CET-6), mainly provide scores or rankings. These abstract results fail to offer detailed insights into students' strengths and weaknesses in reading comprehension. As a result, they do not fully serve the purpose of assessment or meet students' practical needs. With the advancement of information technology, the application of computer-based approaches in education has gained growing attention. Cognitive diagnostic models, which provide detailed and fine-grained analysis, align well with Chinese students' needs for personalized feedback. This study adopts the Generalized DINA (G-DINA) model to analyze students' English reading performance and generate personalized diagnostic feedback reports. The results show that the G-DINA model fits the college English reading test data effectively and yields meaningful diagnostic information. A survey of students' feedback and suggestions reveals that most students responded positively to the content and design of the feedback reports. They believed that personalized feedback enhanced their learning experience and contributed to instructional improvement. By incorporating cognitive diagnostic assessment into English reading instruction, this study proposes a more accurate, data-driven feedback approach. It addresses the limitations of traditional assessments and supports more targeted teaching improvements in English education at Chinese universities.

Keywords: College English Reading; Cognitive Diagnostic Assessment; G-DINA Model; Diagnostic Feedback

*CORRESPONDING AUTHOR:

Zhe Zhang, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; Email: p115768@siswa.ukm.edu.my

ARTICLE INFO

Received: 6 February 2024 | Revised: 25 February 2024 | Accepted: 3 March 2025 | Published Online: 13 March 2025
DOI: https://doi.org/10.30564/fls.v7i3.8674

CITATION

Zhang, Z., Hamzah, M.I., Jamaludin, K.A., 2025. A Framework for Constructing Cognitive Diagnostic Reports on English Reading for Chinese University Students. Forum for Linguistic Studies. 7(3): 643–655. DOI: https://doi.org/10.30564/fls.v7i3.8674

1. Introduction

With the development of the economy and China's increasing involvement in globalization, English continues to play an important role in international economic trade, cultural exchanges, and academic collaboration. To meet the needs of society and the market, China's higher education system has gradually recognized the importance of English education and feedback for non-English majors. Strong English proficiency significantly impacts the academic performance and career prospects of Chinese university students[1]. English education is not only crucial for promoting the internationalization of higher education but also plays an important role in enhancing the global competitiveness of China's workforce[2]. In response, China's Ministry of Education released an updated version of the "Teaching Requirements for College English Programs" in 2019, emphasizing the need to develop students' ability to use English effectively in academic, professional, and social settings[3]. To achieve this goal, recent teaching reforms have placed greater emphasis on enhancing students' productive skills, such as listening and speaking, while also strengthening their reading proficiency[4]. As a key form of language input, reading accounts for over 90% of total language exposure, providing the foundation for developing listening, speaking, and writing skills. Only when students reach a certain level of reading proficiency can they achieve significant improvements in other language abilities[5]. Despite these efforts, many college students still struggle with reading. The College English Test-Band 4 (CET-4), a standardized exam required for non-English majors, is a graduation requirement at many universities[6]. In 2022, approximately 600,000 students took the CET-4, where the reading section accounted for 35% of the total score, or 248 points[7]. However, the average reading score was only 135, falling below the commonly accepted passing threshold of 60%. According to "China's Standards of English Language Ability", a national framework categorizing English proficiency into nine levels, the average CET-4 reading score was significantly below Level 5, the minimum level expected for university students[8].

Several key challenges currently hinder the development of English reading education in Chinese universities. First, the heavy reliance on summative assessments means students typically receive only an overall score, with no detailed feedback on individual questions or instructor comments.The abstract score-based assessment does not allow teachers to quickly identify students' strengths and weaknesses in different reading skills, making it difficult to adjust teaching content and provide targeted remedial instruction. While summative assessment has advantages in statistical scoring, horizontal comparison of student performance, and student selection, it is not conducive to personalized teaching. Moreover, score-based assessment alone does not help students understand the details of reading skills in a timely manner, limiting their ability to learn independently[9].

Currently, the assessment model in China primarily focuses on selection and ranking. To pass exams, both students and teachers prioritize test preparation over improving actual English proficiency. This has led to a disconnect between university English teaching theories and practical instruction, with test-taking strategies becoming the focus of many non-English major courses. Cognitive Diagnostic Assessment (CDA) and cognitive diagnostic reports provide an effective solution for improving students' reading abilities and practical English skills[10]. Diagnostic reports, which are based on students' individual reading characteristics, offer detailed analyses of their performance in different reading skills. This helps both teachers and students gain a clearer understanding of their learning progress and needs. By integrating students' overall reading ability with their mastery of specific reading skills, such feedback provides valuable information for personalized teaching, thereby improving classroom efficiency and learning outcomes[10, 11].

This study aims to analyze university students' English reading ability using the G-DINA cognitive diagnostic model and generate personalized diagnostic feedback reports. Research and application of English reading cognitive diagnostic reports are crucial for enhancing Chinese university students' reading abilities. By refining and improving these diagnostic tools, educators and researchers can provide more targeted support for personalized teaching, helping students effectively overcome reading difficulties and improve their overall English proficiency. The research questions are as follows:

(1) How can personalized English reading feedback reports be designed and generated?
(2) How can the effectiveness of personalized English reading feedback reports be evaluated?

2. Literature Review

2.1. Cognitive Diagnostic Assessment (CDA)

Cognitive Diagnostic Assessment (CDA) is based on the cognitive diagnostic theory and gradually gained attention in the 1990s. CDA measures students' knowledge acquisition and cognitive abilities through cognitive diagnostic models, providing detailed information about their cognitive strengths and weaknesses[12]. This information is presented to teachers and students through cognitive diagnostic feedback (DF) reports. The diagnostic results not only include the student's ability and cognitive level, but also distinguish differences in cognitive abilities among students with similar test scores. For teachers, the detailed student test information improves the efficiency of teachers in refining their teaching content, making their teaching goals more relevant and increasing the significance of remedial teaching. For students, getting detailed test feedback in time after the test can clarify the strengths and weaknesses embodied in the test for individuals, providing a clear direction for themselves to complete remedial learning, and facilitating the improvement of self-learning ability. At the same time, the personalized advice of the test results makes students no longer focus only on scores and rankings, but more on how to improve and perfect, which mobilizes students' learning motivation and is in line with the goals of college English teaching[13, 14].

Since Nichols proposed a framework for CDA development[15], various cognitive diagnostic frameworks and methods have been established to provide diagnostic information for classroom teaching and learning[16–18]. While the specific steps for adapting existing tests into CDA may vary, the process generally includes the following:

(1) Defining cognitive attributes: Cognitive attributes are the specific knowledge, skills, or strategies required to complete test tasks. In reading comprehension, typical attributes include identifying main ideas, recognizing details, and making inferences. Attribute definitions can be based on test specifications, content domain theories, item content analysis, and think-aloud protocols[19, 20].

(2) Constructing the Q-matrix: Tatsuoka[21, 22], a pioneer in Q-matrix theory, proposed that the Q-matrix detects learners' latent knowledge states and represents abstract knowledge through observable theoretical models. The Q-matrix links test items to cognitive attributes, where "0" indicates that answering the item correctly does not require the attribute, and "1" indicates that the attribute is necessary to answer the item correctly[23]. The Q-matrix is a critical input for applying Cognitive Diagnostic Models (CDMs) to generate diagnostic data on students' skill mastery.

(3) Data analysis: The constructed Q-matrix is analyzed using CDMs, grouping students based on their mastery of attributes. Diagnostic information is then provided to teachers or students through score reports or diagnostic summaries. If the model's data fit is insufficient, the attributes and Q-matrix should be adjusted until appropriate results are obtained. This process provides diagnostic insights at both the individual and group levels[22].

(4) Score reporting/Diagnostic Feedback: The ultimate goal of CDA is to provide test-takers with information on their skill mastery and offer improvement strategies through diagnostic feedback (DF)[24]. DF helps students understand their strengths and weaknesses, while teachers can use the feedback to refine instruction, create personalized learning plans, recommend independent study strategies, and improve course design[25]. However, it is important to note that DF reports alone do not directly improve teaching and learning. Their effectiveness depends on how users perceive their quality and usefulness, which motivates them to apply the feedback for skill development[14].

2.2. Cognitive Diagnostic Models (CDMs)

In recent years, cognitive psychologists and psychometricians have developed various psychometric models with diagnostic functions, known as Cognitive Diagnostic Models (CDMs). CDA applies CDMs to analyze students' responses, providing insights into their cognitive strengths and weaknesses, which can help improve teaching and learning strategies[23]. CDMs are latent variable models primarily used to assess students' mastery of multiple skills. They integrate cognitive psychology, item response theory, and statistical modeling. Based on their assumptions, CDMs can be classified into non-compensatory, compensatory, and general models[26, 27]. For example, in an English reading test that requires students to master three reading skills, non-

compensatory models (such as the Fusion Model and Attribute Hierarchy Model) assume that all three skills must be mastered for success. In contrast, compensatory models (such as G-DINA) assume that well-developed skills can compensate for weaker ones. However, due to the complexity of certain domains, researchers may find it challenging to determine whether skills are compensatory.

Selecting an inappropriate model can lead to inaccurate diagnoses. G-DINA (Generalized Deterministic Input, Noisy "And" Gate Model, G-DINA) is one of the cognitive diagnostic models that has been applied to the field of education more times than others.The DINA model is a classic psychometric model that is mostly used to diagnose the test taker's mastery of specific cognitive skills [28]. It combines conjunctive and noisy characteristics to evaluate skill mastery across multiple areas, de la Torre further developed the G-DINA model in 2011 to make it more widely applicable. Unlike the DINA model [29], G-DINA does not assume that testers who do not master certain attributes have the same probability of answering correctly. It uses an analysis of variance (ANOVA)-like modeling approach to analyze data through main and interaction effects. In addition, more specific cognitive diagnostic models can be derived from the G-DINA model by removing specific main or interaction effects.The ability of the G-DINA model to accommodate both compensatory and non-compensatory cognitive attributes makes it more flexible and broadly applicable than other models. Language tests are abstract and diverse, and the G-DINA model's combination of flexibility and compensatory features makes it more popular for reading skill tests, and thus has become the first choice of many researchers in recent years. For example, Chen & Chen applied the G-DINA model to the PISA English reading test by analyzing the response data of 1,029 test takers [30], and the results showed that the model was able to efficiently capture five expert-defined attributes of reading comprehension. Boori et al.applied the G-DINA model to the IELTS reading test data by optimizing the initial Q matrix to estimate the probability of mastery of cognitive attributes for 1,025 test takers [31]. Hemmat et al. used the G-DINA model to analyze the reading comprehension section of the Iranian National University Entrance Examination and found that about 57% of the students had not yet mastered the required reading attributes [32]. The results of their study provide valuable diagnostic feedback for English as a Foreign Language (EFL) teachers to help them gain a deeper understanding of their students' reading ability

2.3. Previous Research

Over the past three decades, cognitive diagnostic theory has gained significant attention in psychometrics and educational research. Researchers have integrated it into large-scale testing, leading to internationally recognized diagnostic testing systems such as the Diagnostic English Language Needs Assessment (DELNA) in Australia, the European DIALANG system, the Finnish second/foreign language reading and writing diagnostic system (DIALUKI), and Hong Kong's Diagnostic English Language Tracking Assessment (DELTA). These systems apply cognitive diagnostic theory across multiple dimensions, enhancing the accuracy and relevance of language assessment. However, international research on cognitive diagnostic testing has largely focused on standardized English proficiency tests, with limited studies on school-based assessments that directly impact learning and teaching.

In China, English proficiency assessment relies primarily on exam scores, such as university-based tests and national exams like CET-4 and CET-6. Although the College English Test (CET-4 and CET-6), as a national-level English test, can objectively and comprehensively assess students' English proficiency, the feedback provided by such exams is limited to abstract scores. Students cannot fully understand their strengths and weaknesses just through scores, and the individual differences among students with the same score cannot be ignored [33, 34]. Moreover, this approach fosters a test-oriented mindset rather than long-term language development. Despite the increasing emphasis on English education in China, challenges remain, particularly in university English reading instruction. Chinese researchers have increasingly explored cognitive diagnostic assessment (CDA) as a means to enhance reading assessment.

While previous studies have validated cognitive diagnostic models and identified reading attributes, fewer studies have integrated diagnostic results into personalized feedback. This study aims to address this gap by exploring how diagnostic feedback can effectively support student learning and instructional decision-making. Chinese researchers typically follow a similar approach when integrating cognitive diagnostic theory with school-based assessments. They first se-

lect a cognitive diagnostic model, then determine the reading attributes assessed in school-based tests through think-aloud protocols and expert judgment, and finally apply the model to diagnose students' mastery of these attributes [11, 18, 35–40].

Jang employed the Fusion Model to diagnose cognitive attributes in students' reading test responses using verbal reports [18]. Through a mixed-methods approach, Jang identified nine reading ability parameters. His diagnostic research on fusion modeling has since gained widespread acceptance and has become a research focus. Wang and Gier applied the Attribute Hierarchy Model to diagnose and analyze students' cognitive skills in SAT critical reading [35]. Their study verified the validity of attribute hierarchies through qualitative and quantitative research using think-aloud protocols and hierarchical consistency indicators, ultimately categorizing students' attribute mastery levels. Du and Ma constructed a cognitive diagnostic model for English reading based on cognitive diagnostic assessment and validated reading cognitive attributes through tree regression analysis [36], identifying eight distinct reading attributes. While previous reading studies have used various diagnostic models and methods with promising results, research has primarily focused on constructing and validating reading cognitive attributes, with limited emphasis on evaluating the effectiveness of diagnostic feedback.

The studies mentioned above indicate that cognitive diagnostic research in China has rarely focused on feedback reports. Currently, there is a lack of standardized criteria for high-quality DF reports, and limited research has explored their effectiveness or practical application. DF reports are constructed based on students' reading test responses and cognitive diagnostic assessments, incorporating key features in both content and presentation. Henderson emphasized that DF reports should present test content, score interpretations [41], and diagnostic information in clear and concise language to ensure that learners accurately understand the feedback and adjust their learning accordingly. The content and presentation of the DF report should be scientifically sound and easy to understand. In terms of content, the report should contain scores, diagnostic results, performance analysis and learning suggestions [42]. Presenting traditional scores in conjunction with cognitive diagnostic results helps teachers and students understand that the same test scores may correspond to different patterns of mastery of cognitive attributes [43]. Combining the analysis of student performance with remedial learning or instructional recommendations helps report recipients to reflect on past learning and develop more efficient learning strategies, thus fulfilling the core goal of feedback [36]. In terms of presentation, the order of the report content is very important; presenting student performance first and then analyzing the cognitive diagnosis results can reduce students' rejection and confusion about professional analysis [42]. DF reports should be individualized to meet the specific needs of learners [43]. The design of the DF report in this study was based on the research of [44]. They suggested that personalized feedback reports can increase students' interest in performance feedback and reduce learners' fear and stress about testing. Therefore, personalized feedback on the mastery of reading attributes was included in the student feedback reports to help students gain a clearer understanding of their strengths and weaknesses and develop a targeted improvement plan.

3. Research Methods

3.1. Participants

This study involved 200 students in the questionnaire survey phase. To ensure representativeness, a random sampling method was used, giving each student an equal chance of being selected. In the think-aloud experiment phase, five student volunteers participated, all of whom were non-English majors with intermediate to advanced English proficiency. Participants were selected through purposive sampling based on their academic performance, specifically those ranked in the top one-third of their class, and recommendations from their English teachers to ensure they met the specific criteria required for the study.

Additionally, five university English teachers with over ten years of teaching experience were invited as experts to identify key cognitive attributes and learning strategies in English reading. All experts were from the same university where the study was conducted and had professional expertise in English reading, ensuring the relevance and credibility of their insights.

3.2. Research Instruments

The reading comprehension test in this study was based on the College English Test Band 4 (CET-4). The CET-4 is

designed by a national expert team according to the exam syllabus and university students' actual English needs, making it a reliable and standardized assessment tool[45]. The test consisted of three passages: the first passage included 10 information-matching questions, while the other two passages each contained 5 multiple-choice questions, totaling 20 questions. Each question was worth 1.5 points, and all were scored using a dichotomous rating system. The test covered two types of texts: expository and argumentative essays. The first passage was approximately 600 words, while the other two were around 350 words each.

The researcher analyzed the test items and results from four dimensions: difficulty, discrepancy, guessing coefficient, and slip coefficient. Overall, the difficulty of the items was moderate, with an average difficulty of 0.56, making it suitable for assessing students of different ability levels. In terms of discrepancy, the average discrepancy value for all items was 0.48, indicating that there was considerable variation in student performance on most items, which may be due to higher difficulty or differences in how students understood the items. For example, Item 3 (discrepancy 0.50) had a higher discrepancy, indicating that students' answers were more spread out, which could be related to the design of the question or the ability factors being assessed. In contrast, Item 11 (discrepancy 0.44) showed less variation, meaning most students were able to answer the question correctly. All items had a guessing coefficient of 0.33 because each item was a multiple-choice question with four options, so the probability of guessing the correct answer was 1/3. Therefore, guessing does not reflect the students' true abilities. The average slip coefficient was 0.10, indicating that there was a certain degree of careless errors across the items. Some items, like Item 5 (slip coefficient 0.1462) and Item 14 (slip coefficient 0.1391), had higher slip coefficients, which may be due to ambiguities in the design of these items, leading to more mistakes by students. Overall, the test items were designed at an appropriate level and could effectively distinguish between students of varying abilities, but there is still room for improvement in the design of certain items, such as those with high slip coefficients.

A panel of English assessment experts reviewed the test content to ensure its validity in measuring students' reading ability and its alignment with instructional objectives. To assess the reliability of the test, a test-retest method was employed, where the same group of students took the test twice with a one-month interval. Pearson correlation analysis was conducted to examine the relationship between the two test scores. The results showed a correlation coefficient of $r = 0.75$ ($p < 0.001$), indicating a strong positive correlation. This suggests that the test demonstrated high reliability and effectively measured students' English reading proficiency.

To explore students' perceptions of the diagnostic feedback report, data were collected through a questionnaire adapted from Zhang Haiyun[38]. The questionnaire consisted of four sections, covering demographic information, evaluation of the design and content, acceptance and willingness to use, and open-ended feedback. The design and content evaluation focused on assessing the clarity, structure, and usefulness of the diagnostic feedback report, while the acceptance and willingness section measured students' attitudes toward the report and the likelihood of using it for learning improvement. These two sections included a total of ten Likert-scale multiple-choice questions. To evaluate the internal consistency of these items, Cronbach's α coefficient was calculated, yielding a result of 0.85, which indicates strong reliability. This suggests that the questionnaire provided a stable and consistent measure of students' evaluations of the diagnostic feedback report.

3.3. Data Collection and Analysis

This study applied cognitive diagnostic assessment (CDA) theory to the evaluation and feedback of English reading performance. The research process involved defining reading attributes, constructing a Q-matrix, conducting cognitive diagnostic analysis, and generating diagnostic feedback reports. The reading attributes were determined through think-aloud protocols and expert discussions. Five reading experts analyzed and discussed the cognitive attributes involved in the test items. In cases of disagreement, only attributes recognized by at least three experts were retained. When multiple attributes were associated with a single item, they were ranked based on their importance in the problem-solving process.

Through this process, eight key cognitive attributes in reading were identified: vocabulary recognition, understanding literal meaning at the sentence level, understanding literal meaning at the passage level, inferring word meaning, constructing textual cohesion inferences, making text-based

elaborative inferences, information filtering, and information summarization. This structured approach ensured that the cognitive diagnostic assessment and feedback reports were both precise and effective. The expert panel then refined these attributes further, merging those that appeared infrequently or overlapped across multiple items. This led to the final selection of eight reading attributes, as shown in **Table 1**. The initial Q-matrix was then constructed and verified using the G-DINA model. Based on the verification results, the expert panel conducted additional analyses and refinements, producing the final Q-matrix presented in **Table 2**. Fleiss's Kappa index is a common metric for measuring inter-rater agreement among multiple assessors. This study employs this index to estimate the consistency of the expert group's labeling outcomes. According to the standards set by Landis and Koch (1977), a Kappa value between (0.21, 0.40) is considered "acceptable," between (0.41, 0.60) indicates "moderate" consistency, between (0.61, 0.80) reflects "good" consistency, and a Kappa value greater than 0.81 is deemed "almost perfect." The five experts labeled 20 reading items in detail, aiming to identify the necessary attributes for correctly answering each item. The Kappa values indicated that the experts' judgment consistency on eight attributes was 0.43, 0.52, 0.55, 0.38, 0.44, 0.51, 0.49 and 0.47, with seven values falling within the range of (0.4, 0.6), except for the fourth, which was slightly below 0.4. The Kappa values generally indicate good consistency, suggesting the Q-matrix constructed by the expert group is fundamentally valid.

Table 1. Attributes.

Ability	Attributes	Description
Recognition	A1. Recognizing General and More Difficult Vocabulary.	Understanding general vocabulary/phrases in questions, options, or target sentences. Understanding specialized or more challenging vocabulary/phrases in questions, options, or target sentences.
Comprehension	A2. Understanding Sentence Meaning.	Processing and understanding complex and lengthy sentences, especially those that are critical to comprehension.
	A3. Understanding the Literal Meaning of Text.	Understanding the literal meaning of multiple sentences, including paraphrasing.
Inference	A4. Inferring Vocabulary from Context.	Inferring the meaning of vocabulary/phrases based on context.
	A5. Inferring Indirect Information from Text.	Understanding and inferring the implicit meaning of sentences, paragraphs, or the entire text.
	A6. Inferring Pragmatic Meaning from Text.	Making inferences based on sociolinguistic and sociocultural knowledge.
Analysis	A7. Distinguishing Relevant and Irrelevant Information	Differentiating between relevant (or important) and irrelevant (or unimportant) information, then focusing on the relevant or important information.
	A8. Analyzing the Author's Viewpoint/Intention and Summarizing the Passage Theme	Understanding the author's viewpoint, bias, values, or intentions, and summarizing the theme of a paragraph or passage.

Table 2. Q-matrix.

Attributes	A1	A2	A3	A4	A5	A6	A7	A8
1	1	1	0	0	0	0	0	1
2	1	0	1	0	0	1	0	0
3	0	1	0	1	0	0	0	0
4	1	0	0	0	1	0	1	0
5	0	0	1	1	0	0	1	0
6	0	1	0	0	1	0	0	1
7	0	0	1	0	0	1	1	0
8	0	0	0	1	0	0	0	1
9	1	0	0	0	0	1	0	0
10	0	0	1	0	0	0	0	1
11	0	0	0	1	1	0	0	0
12	0	0	0	0	0	1	1	0
13	0	1	0	0	1	0	0	0
14	0	0	0	0	0	1	0	1
15	1	0	1	0	0	0	0	0
16	0	1	0	0	0	1	0	1
17	0	0	0	1	0	0	0	0
18	1	0	0	0	1	0	1	0
19	0	0	1	1	0	0	0	1
20	0	1	0	0	0	1	0	0

The G-DINA model was applied for cognitive diagnostic analysis, and the model's fit with the data served as the basis for generating the DF reports. To ensure students could effectively interpret the reports, all instructors received training on how to explain them. They were then responsible for distributing and clarifying the DF reports to the students.

3.4. Questionnaire Data Collection and Analysis for DF Report Evaluation

The questionnaire was distributed through *Wenjuanxing* (an online survey platform) for students to complete after reviewing their DF reports. The data analysis focused on three key aspects: students' satisfaction with the content and design of the DF reports, their acceptance of specific elements within the reports, and their suggestions for further improvements.

4. Findings

4.1. Student Diagnostic Score Report

A total of 200 students participated in the reading test, with an average score of 23.37 out of 30. The median score was 25, and the standard deviation was 6.01. The student diagnostic score report consists of four sections. The first section presents the attribute mastery levels (**Figure 1**) and an analysis of students' relative strengths and weaknesses, providing an overall assessment of their reading ability. By reviewing the figure and the summary of their strengths and areas for improvement, students can clearly identify their reading skills that need further development.

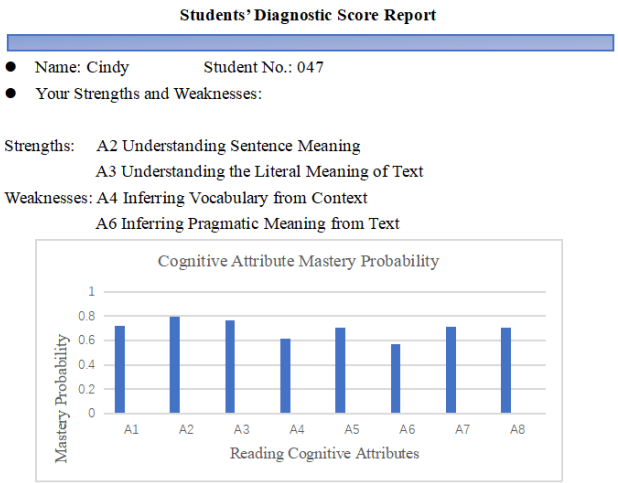

Figure 1. Reading cognitive attributes.

The second section, "Question-Level Feedback (**Table 3**)," provides test-takers with correct answers to the test questions, their total score, and their ranking within the class. This section offers objective information about the test performance. By reviewing their responses, test-takers can easily identify which questions were answered correctly and which were incorrect. This helps them understand their mistakes and focus on areas that need improvement. Additionally, the overall score and class ranking provide insight into their relative proficiency in English, aiding in goal setting for future improvement. This feedback not only motivates test-takers to study diligently but also ensures their privacy.

Table 3. Question-level feedback.

Items	Your Answer	Correct Answer
1	✓	A
2	✓	C
3	✓	D
4	✓	A
5	✓	B
6	✓	C
7	✓	D
8	B	C
9	D	B
10	✓	A
11	A	D
12	C	B
13	✓	C
14	D	A
15	✓	D
16	✓	C
17	C	A
18	✓	B
19	✓	D
20	✓	C
Your total score: 14 (20)		
You scored equal to or higher than 51% of students.		

The third section, "Primary Attribute/Skill Descriptions and Sample Questions" (**Table 4**), provides detailed descriptions of eight reading attributes/skills, along with the questions associated with each. This section aims to help both teachers and students review the test. The descriptions offer students a clear learning framework, enabling them to understand various reading skills and attributes. This helps students identify specific areas to focus on during their reading practice and provides a standard for assessing their reading abilities. Additionally, the descriptions allow students to recognize their strengths and weaknesses, guiding them toward targeted practice and improvement to enhance overall reading proficiency.

Table 4. Primary attribute/skill descriptions and sample questions.

	Attribute/Skill Descriptions	Example Questions
A1	**Recognizing General and More Difficult Vocabulary** Understanding general vocabulary/phrases in questions, options, or target sentences. Understanding specialized or more challenging vocabulary/phrases in questions, options, or target sentences.	1.2.4.9.15.18
A2	**Understanding Sentence Meaning** Processing and understanding complex and lengthy sentences, especially those that are critical to comprehension.	1.3.6.13.16.20
A3	**Understanding the Literal Meaning of Text** Understanding the literal meaning of multiple sentences, including paraphrasing.	2.5.7.10.15.19
A4	**Inferring Vocabulary from Context** Inferring the meaning of vocabulary/phrases based on context.	3.5.8.11.17.19
A5	**Inferring Indirect Information from Text** Understanding and inferring the implicit meaning of sentences, paragraphs, or the entire text.	4.6.11.13.18
A6	**Inferring Pragmatic Meaning from Text** Making inferences based on sociolinguistic and sociocultural knowledge.	2.7.9.12.14.16.20
A7	**Distinguishing Relevant and Irrelevant Information** Differentiating between relevant (or important) and irrelevant (or unimportant) information, then focusing on the relevant or important information.	4.5.7.12.18
A8	**Analyzing Author's Viewpoint/Intention and Summarizing the Passage Theme** Understanding the author's viewpoint, bias, values, or intentions, and summarizing the theme of a paragraph or passage.	1.6.8.10.14.16.19

The final section, "Suggestions for Learning," provides recommendations for further learning based on the student's strengths and weaknesses. An example based on the findings is provided below:

> *"When reading argumentative and expository passages at the CET-4 level, you excel at summarizing main ideas and decoding explicit information. However, your skills in understanding lexical meaning and analyzing syntactic structure need improvement. Expanding your vocabulary and mastering syntactic knowledge will help in these areas."*

Targeted learning suggestions offer personalized instruction that effectively addresses students' individual needs. By focusing on areas for improvement, students can avoid spending time on content they have already mastered, thereby increasing learning efficiency. Specific recommendations also serve as motivation. As students make progress based on these suggestions, their confidence grows, enabling them to approach learning challenges with a more positive attitude.

4.2. Evaluation of the DF Report

This survey aimed to gather deeper insights into students' attitudes toward the DF report and their approval of its content. A total of 189 students who received the DF report completed the questionnaire, resulting in a valid response rate of 94.5%. The content and design of the DF report covered four areas: mastery of reading attributes, analysis of the strengths and weaknesses of these attributes, feedback on test answers and personal scores, and the classification and analysis of reading attributes in relation to test questions, along with personal suggestions (**Table 5**). Over 91% of students reported that they liked or very much liked the DF report overall. Key elements that should be emphasized include mastery of reading attributes (85%), classification and analysis of reading attributes with test questions (89%), and personal suggestions (92%).

The approval of the DF report focused on students' recognition of their mastery of eight reading attributes, their willingness to improve reading skills based on the provided learning suggestions, and the report's overall usefulness for their English learning (**Table 6**). Regarding the recognition of mastery for the eight reading attributes, over 82% of students agreed with the assessment of their proficiency. Additionally, more than 86% of respondents believed that the DF report helped them understand their strengths and weaknesses, supporting their continued improvement.

The survey concluded with two open-ended questions. The first asked students to provide their overall impression of the DF report and list its strengths and weaknesses. Most students provided highly positive feedback, frequently using terms such as "excellent," "clear," "intuitive," and "visually appealing." Many also expressed their intention to follow the

remedial learning suggestions to improve their academic performance. However, some students requested more detailed explanations of the report's content. The second open-ended question asked, "Apart from the current information presented in the report, what additional content would you like to see included?" In response, most students felt that the existing content was sufficient. Some suggested adding a comprehensive analysis of skills such as listening, writing, and translation, while a few recommended including explanations for incorrectly answered questions.

Table 5. Areas of DF report.

Quality	Very Much Dislike	Dislike	Neutral	Like	Very Like
1. The Mastery of Reading Attributes	7 (3.70%)	8 (4.23%)	13 (6.87%)	67 (35.45%)	94 (49.73%)
2. Strengths and Weaknesses of Attributes	11 (5.82%)	15 (7.94%)	45 (23.81%)	59 (31.22%)	59 (31.22%)
3. Test Answers and Personal Scores	8 (4.23%)	14 (7.41%)	65 (34.39%)	69 (36.51%)	33 (17.46%)
4. Primary Attribute/Skill Descriptions and Sample Questions	5 (2.6%)	6 (3.17%)	9 (4.76%)	80 (42.33%)	89 (47.09%)
5. Personal Suggestions	3 (1.59%)	6 (3.17%)	6 (3.17%)	95 (50.26%)	79 (41.79%)
6. Overall Assessment	3 (1.59%)	5 (2.65%)	9 (4.76%)	84 (44.44%)	88 (46.56%)

Table 6. Acceptance of the DF report.

Acceptance	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Assessment of Mastery of Reading Attributes	6 (3.17%)	7 (3.7%)	21 (11.11%)	57 (30.16%)	98 (51.85%)
Plan to Improve Reading Skills Based on Learning Suggestions	13 (6.88%)	17 (8.99%)	19 (10.05%)	64 (33.86%)	76 (40.2%)
The report is very helpful for English learning.	3 (1.59%)	9 (4.76%)	14 (7.41%)	79 (41.79%)	84 (44.44%)

5. Discussion

In this study, a cognitive diagnostic model was applied to college English reading performance to generate a cognitive model of reading ability. The model's definition and categorization of attributes are based on the Curriculum Requirements, which are more suitable for the actual reading status of Chinese college students and can more accurately reflect the cognitive process of reading comprehension. Compared with the existing literature, the model focuses on the refinement of the definition of reading attributes, which helps to determine and label the attributes at a later stage and guides the development of cognitive diagnostic tests. It not only assesses students' language ability, but also explores the developmental characteristics of students' cognitive ability in the process of language learning, which is more comprehensive and specific than the existing models, providing an interpretable theoretical basis for cognitive diagnostic analysis and an operable practical guidance for subsequent remedial teaching.

According to the results of the questionnaire survey, the cognitive diagnostic feedback report (DF) has improved students' understanding of their own reading ability to a certain extent, and is conducive to students' identification of their learning strengths and weaknesses. Moreover, most of the students thought that the personalized feedback provided by the report was important for them to adjust their learning strategies. Students can learn about their mastery of vocabulary, information extraction, and reasoning skills through the report and choose more targeted practice accordingly. This finding is consistent with constructivist learning theory. This theory suggests that learners are able to actively construct knowledge as they interact with their environment, and that personalized feedback is an important tool for promoting self-directed learning. In addition, the visual data (e.g., bar graphs) reported by DF can intuitively show the degree of mastery of different reading ability attributes, enabling students to more clearly understand their strengths and weaknesses in the reading process. This is consistent with the findings of[35, 36, 38–40], who noted that combining traditional

scores with cognitive diagnostic results can be more effective in helping students understand the different cognitive characteristics behind the same scores. The questionnaire showed that most students had a positive attitude towards the DF report, especially the remedial learning advice section, which respondents generally believed could provide specific guidance for their subsequent learning. For example, 65% of the students indicated that they would adjust their learning strategies according to the recommendations listed in the report, such as increasing practice on reasoning topics or strengthening training on information extraction skills. This finding is consistent with the findings of Alamri et al. [46]. This study indicated that personalized learning feedback can increase students' motivation and focus on future learning.

Despite the fact that the DF reports provide detailed cognitive diagnostic data, especially when the reports involve more complex cognitive diagnostic models, some students still encounter comprehension difficulties when reading the reports. Some students reported that some of the terms used in the report were difficult to understand, such as "probability of mastery of cognitive attributes", which was an abstract expression that hindered their overall interpretation of the report. Some respondents also suggested that the usefulness of the report could be further enhanced if more intuitive explanatory notes were included in the report, or a brief reading guide was provided. For example, Zhang and Hyland pointed out that the readability of test feedback is closely related to its practical application [42], and optimizing the report language can effectively improve the usability of the feedback. Therefore, future feedback report design could consider using more concise terminology and incorporating case examples to explain student learning to lower the threshold of understanding.

In addition, although the DF reports in this study were able to provide personalized feedback for different students' reading abilities, certain limitations still existed. For example, it was found that some students would like the report to provide more specific learning suggestions, such as recommending appropriate reading materials or practice topics. The current feedback report mainly focuses on the diagnosis of reading ability, and is not yet capable of recommending targeted learning resources. Subsequent research could further incorporate computer technology to match students with learning content that better meets their individual needs based on their cognitive diagnostic data, in order to enhance the personalization of the feedback.

This study was conducted with sophomore non-English majors at a university in China; therefore, the applicability of the findings to different learning groups needs to be further verified. For example, more data are needed to support whether students in different grades and with different levels of English proficiency will have different acceptance of DF reports. In addition, the data collection in this study mainly relied on questionnaires and failed to incorporate interviews or behavioral observations, which may lead to limitations in some of the feedback.

6. Conclusions

This study applied the G-DINA model to assess Chinese university students' English reading proficiency, focusing on their mastery of eight cognitive attributes and subsequently provided students with diagnostic feedback (DF) reports consisting of four sections. The study also examined students' perceptions and evaluations of the DF reports. While overall feedback was positive, some limitations remain. First, the time required to generate the DF reports resulted in a one-month delay in delivery, reducing the immediacy and potential impact of the feedback. Additionally, despite efforts to present clear and concise remedial recommendations, some students may still find them difficult to interpret, highlighting the need for supplementary support materials and targeted exercises from educators and researchers. Future research should explore how both teachers and students utilize the DF reports and assess their effectiveness in improving students' English reading skills.

Author Contributions

All authors have made a substantial, direct, and intellectual contribution to the work, including but not limited to Conceptualization, Methodology, Investigation, Formal analysis, Writing—Original Draft, and Writing—Review & Editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

All participants provided informed consent before participating in the study. The anonymity and confidentiality of the participants were guaranteed, and participation was completely voluntary.

Data Availability Statement

Data will be made available on request.

Acknowledgments

The authors appreciate the editors and all reviewers for their comments and suggestions. The authors also thank you the people who patiently participated in the study.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Baoqi, S., Hu, G., Curdt-Christiansen, X.L., 2020. Metalinguistic contribution to reading comprehension: A comparison of Primary 3 students from China and Singapore. Applied Psycholinguistics. 41(3), 657–684.

[2] Liu, J.D., Han, B.C., 2018. Theoretical considerations for developing use-oriented Chin's Standards of English. Modern Foreign Languages. 41(1), 78–90.

[3] Zeng, M., Bao, Z.J. 2022. The Dilemma of College English Education and Teaching and Innovative Practices under the Guidance of College English Teaching Guide (2020 Edition). Journal of Jilin College of Education. 38(10), 26–30.

[4] Yanwei, J., 2019. Action Research as a Model for the Enhancement of Communicative Competence among Learners of English as a Foreign Language in Mainland China. In Proceedings of the 2019 3rd International Conference on Education Technology and Economic Management; Xi'an, China, 1–5 September 2019. pp. 909–913.

[5] Zhang, H., Shi, Y., 2020. The group and individualized diagnosis on EFL reading in university-based final test. Technology Enhanced Foreign Language Education. 2020(5), 79–85.

[6] Zheng, R., 2023. Chinese university students' English reading strategies: An analysis of the current situation and a proposal for enhancement. Pacific International Journal. 6(2), 226–231.

[7] Lin, C., Shao, Z.C., Hui, G., 2023. Research and analysis of factors influencing the passing rate of CET-4 in local universities. Lavai-International Journal of Education. 19(29), 30–41.

[8] Han, F., 2021. Washback of the reformed college English test band 4 (CET-4) in English learning and teaching in China, and possible solutions. In: Lanteigne, B., Coombe, C., Brown, J.D. (eds.). Challenges in Language Testing Around the World: Insights for language test users. Springer: Singapore. pp. 35–46.

[9] Zhou, S., McKinley, J., Rose, H., et al., 2022. English medium higher education in China: Challenges and ELT support. English Language Teaching Journal. 76(2), 261–271.

[10] Najarzadegan, S., 2022. The effect of learning a CDA model on promoting EFL undergraduates' reading comprehension ability across different proficiency levels. Journal of Modern Research in English Language Studies. 9(4), 75–92.

[11] Wu, Y., 2018. The current situation and innovative reform path of university English teaching under the background of "Internet+". Educational Theory and Practice. 6(2), 57–58.

[12] Leighton, J.P., Gierl, M.J., 2007. Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. Educational Measurement: Issues and Practice. 26(2), 3–16.

[13] Alderson, J.C., 2017. The challenge of (diagnostic) testing: Do we know what we are measuring. In: Fox, J., Wesche, M., Bayliss, D., et al. (eds.). Language Testing Reconsidered. University of Ottawa Press: Ottawa, Canada. pp. 21–39.

[14] Kunnan, A.J., Jang, E.E., 2009. Diagnostic feedback in language assessment. In: Michael, H., Catherine, L., Doughty, J. (eds.). The Handbook of Language Teaching. Wiley-Blackwell Publishing Ltd.: Malden, USA. pp. 610–627.

[15] Nichols, P.D., 1994. A framework for developing cognitively diagnostic assessments. Review of Educational Research. 64(4), 575–603. DOI: https://doi.org/10.3102/00346543064004575

[16] De la Torre, J., Sorrel, M.A., 2023. Cognitive diagnosis models. In: Ashby, F. G., Colonius, H., Dzhafarov, E.N. (eds.). New handbook of mathematical psychology, Volume 3: Perceptual and cognitive processes. Cambridge University Press: Cambridge, UK. pp. 15–24.

[17] Leighton, J.P., 2022. Cognitive Diagnostic Testing. Routledge: London, UK. pp. 50–65.

[18] 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.

[19] Buck, G., Tatsuoka, K., 1998. Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. Language testing. 15(2), 119–157.

[20] Mei, H., Chen, H., 2024. Cognitive diagnosis in language assessment: A thematic review. RELC Journal. 55(2), 529–537.

[21] Tatsuoka, K.K., 2009. Cognitive Assessment: An Introduction to the Rule Space Method. Routledge: New York, NY, USA. pp. 80–98.

[22] Tatsuoka, K.K., 2013. Toward an integration of item-response theory and cognitive error diagnosis. In: Frederiksen, N., Glaser, R., Lesgold, A., et al. (eds.). Diagnostic Monitoring of Skill and Knowledge Acquisition. Routledge: Hillsdale, NJ, USA. pp. 453–488.

[23] Rupp, A., Templin, J., Henson, R., 2010. Diagnostic Measurement: Theory, Methods, and Applications. The Guilford Press: New York, USA. pp. 102–110.

[24] Wei, J., 2016. Bridging the Gap Between Research and Practice: Construction and Validation of a CDA-informed English Reading Test for China's Twelfth Graders [Doctoral dissertation]. University of Illinois at Urbana-Champaign: Champaign, IL, USA. pp. 45–61.

[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.

[26] Sun, H., Kang, Y., 2022. A review of cognitive diagnosis assessment in language from 2000 to 2021: A visualized analysis with CiteSpace II. Journal of Humanities and Social Sciences Studies. 4(3), 182–191.

[27] Chen, Q.M., Zhang, M.Q., 2010. Development of cognitive diagnosis models and their application methods. Advances in Psychological Science. 18(03), 522.

[28] De La Torre, J., 2011. The generalized DINA model framework. Psychometrika. 76(2), 179–199.

[29] Nájera, P., Abad, F.J., Chiu, C.Y., et al., 2023. The restricted DINA model: A comprehensive cognitive diagnostic model for classroom-level assessments. Journal of Educational and Behavioral Statistics. 48(6), 719–749.

[30] Chen, H., Chen, J., 2016. Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. Language Assessment Quarterly. 13(3), 218–230.

[31] Boori, A.A., Ghazanfari, M., Ghonsooly, B., et al., 2024. Cognitive diagnostic modeling analysis of the reading comprehension section of an Iranian high-stakes language proficiency test. International Journal of Language Testing. 14(1), 17–33.

[32] Hemmat, S.J., Baghaei, P., Bemani, M., 2016. Cognitive diagnostic modeling of L2 reading comprehension ability: Providing feedback on the reading performance of Iranian candidates for the university entrance examination. International Journal of Language Testing. 6(2), 92–100.

[33] Wang, H., 2021. Analysis and prediction of CET4 scores based on data mining algorithm. Complexity. 2021(1), 5577868.

[34] Huang, C., 2023. Assessing reading comprehension in CET-4: A cross-sectional case study. Interactive Learning Environments. 31(10), 7149–7158.

[35] Wang, C., Gierl, M.J., 2011. Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. Journal of Educational Measurement. 48(2), 165–187.

[36] Du, W., Ma, X., 2018. CDA-based diagnostic model construction for EFL reading tests. Foreign Language Teaching and Research. 50(1), 74–88.

[37] Luo, J., 2019. Cognitive Diagnostic Assessment for Reading Comprehension Based on College English Placement Test [Doctoral dissertation]. Hunan University: Changsha, China. pp. 78–90.

[38] Zhang, H.Y., 2022. Application and evaluation of diagnostic feedback report on college English reading–Taking G-DINA model as an example. Foreign language e-Learning. 2, 39–47+117.

[39] Fan, T., Sun, B., Zeng, Y., 2023. A comparative study of different cognitive diagnosis models on English reading tests. China Examinations. (05), 53–61. DOI: http://doi.org/10.19360/j.cnki.11-3303/g4.2023.05.007

[40] Fan, T., Zeng, Y., 2019. A reading ability diagnostic model based on the China English Proficiency Scale. Chinese Foreign Languages. (03), 21–31.

[41] Henderson, M., Ryan, T., Boud, D., et al., 2021. The usefulness of feedback. Active Learning in Higher Education. 22(3), 229–243.

[42] Zhang, Z.V., Hyland, K., 2022. Fostering student engagement with feedback: An integrated approach. Assessing Writing. 51, 100586.

[43] Roberts, M.R., Gierl, M.J., 2010. Developing score reports for cognitive diagnostic assessments. Educational Measurement: Issues and Practice. 29(3), 25–38.

[44] Bernacki, M.L., Greene, M.J., Lobczowski, N.G., 2021. A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose (s)? Educational Psychology Review. 33(4), 1675–1715.

[45] Huang, C., 2023. Assessing reading comprehension in CET-4: A cross-sectional case study. Interactive Learning Environments. 31(10), 7149–7158.

[46] Alamri, H., Lowell, V., Watson, W., et al., 2020. Using personalized learning as an instructional approach to motivate learners in online higher education: Learner self-determination and intrinsic motivation. Journal of Research on Technology in Education. 52(3), 322–352.