LANGUAGE ASSESSMENT QUARTERLY
https://doi.org/10.1080/15434303.2022.2140050

Investigating Second Language (L2) Reading Subskill Associations: A Cognitive Diagnosis Approach

Huilin Chen[a], Yuyang Cai[b], and Jimmy de la Torre[c]

[a]School of Education, Shanghai International Studies University, Shanghai, China; [b]School of Languages, Shanghai University of International Business and Economics, Shanghai, China; [c]University of Hong Kong, Hong Kong, China

ABSTRACT
This study uses a cognitive diagnosis model (CDM) approach to investigate the associations among specific L2 reading subskills. Participants include 1,203 Year-4 English major college students randomly selected from the nationwide test takers of Band 8 of Test for English Majors (TEM8), a large-scale English proficiency test for senior English majors in China. Their English reading was measured using a reading comprehension subtest of the TEM8. Based on the CDM output on latent class size estimates, the chi-square test of independence was used to uncover the associations among reading subskills, and odds ratio estimation was used to determine the strengths of those associations. The CDM output on attribute mastery prevalence was used to establish the stochastic direction of the associations between reading subskills. The study has the following findings: a reading subskill network displaying significant subskill associations together with their strengths and directions can be established through a CDM approach, and the patterns of reading subskill associations based on cognitive levels and local/global comprehension resonate with major reading process models and reflect the hierarchical and compensatory characteristics of reading subskills.

摘要
本研究用认知诊断模型(CDM)研究了二语阅读子技能之间的关系。研究被试为1203位随机抽取的英语专业八级考试(TEM8)的考生。TEM8是一项针对中国大学英语专业四年级学生的大规模英语水平测试，被试的英语阅读能力是通过该测试的阅读理解部分试题来测量的。以认知诊断分析输出的潜在类别比例为统计依据，独立卡方检验可以用来揭示阅读子技能之间是否存在关系，比值比估计可以用来确定这些关系的强度。认知诊断分析输出的属性掌握率则可以用来确立阅读子技能间关系的随机方向。研究发现:认知诊断方法可以揭示阅读子技能间存在的显著关系，并估计这些关系的强度和方向;以认知水平和理解全局性为依据的阅读子技能关系模式体现了阅读过程的主要理论，反映了阅读子技能的层次性和补偿性特征。

Introduction

Theoretical reading models (D. E. Rumelhart, 1977; Goodman, 1967; Gough, 1972; Hoover & Gough, 1990; Kintsch, 1998, 1988; D. Rumelhart, 1980; Smith, 1977) emphasize

CONTACT Jimmy de la Torre j.delatorre@hku.hk Faculty of Education, University of Hong Kong, Pokfulam Road, Hong Kong, China
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15434303.2022.2140050

researchers' attention to the relations among components or subskills of first language (L1) reading comprehension. The discovery of L1 reading subskill associations may be conducive to the development of reading models in general and the improvement of teaching and learning pathways for L1 reading. There is growing interest (e.g., Cai, 2020; Cai & Kunnan, 2019; Song, 2008) in conducting empirical studies to explore how L2 reading subskills are related, for evidence showed that L2 reading subskill relationships could be quite different from L1 reading subskill relationships (Nassaji, 2002, 2007; Stanovich, 1980) and the impact on teaching and learning may also be different. However, much fewer such investigations were conducted employing cognitive diagnosis, which may represent test theory for a new generation of tests (Lohman & Ippel, 1993).

Prevailing measurement theories, such as classical test theory that focuses on the overall "true score," and unidimensional item response theory (IRT) models that focus on the overall ability score, cannot provide information about the mastery of finer-grained subskills. In contrast, cognitive diagnosis pays more attention to the finer-grained subskills instead of the overall ability. The finer-grained information from cognitive diagnosis makes it possible for the associations (the equivalent of "relationships" in statistics) among the subskills to be explored.

Therefore, this study tries to carry out an empirical investigation to assess the capacity of cognitive diagnosis, an emerging measurement theory, in disclosing specific L2 reading subskill associations. The specific associations discovered in this study may reflect the general subskill relationship categories (like the hierarchical, compensatory, and non-compensatory) described in some studies (e.g., Li et al., 2016; Ravand & Robitzsch, 2018; Yi, 2016) on application of cognitive diagnosis to L2 reading. This study attempts to perceive some processes special to L2 reading according to the specific associations among reading subskills.

In this research, we intend to investigate whether specific L2 reading subskill associations can be disclosed using the cognitive diagnosis approach and find out how reading subskills are related, which may have methodological implications on the one hand and theoretical implications on the other.

Literature review

Predictors and subskills of reading comprehension

Bernhardt (2011), Grabe (2009), and Koda (1996, 2005) defined reading as a multicomponent process and identified key predictors of L2 reading. On the basis of previous research, Jeon and Yamashita (2014) found that ten predictors of L2 reading were most valuable for in-depth research because they were found to be very important in explaining L2 reading variance in general. The ten predictors are L2 decoding, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension, L2 phonological awareness, L2 orthographic knowledge, L2 morphological knowledge, L2 listening comprehension, working memory, and metacognition. Among them, the first four predictors are most frequently studied. (Jeon & Yamashita, 2014)

Although those measures play essential roles in predicting the general competence of L2 reading comprehension, L2 reading comprehension needs to be better understood by examining specific comprehension subskills themselves, rather than other language and cognitive constructs that contribute to successful reading. Well-known lists of reading subskills include Davis (1968) eight-subskill classification, Munby's 19-subskill taxonomy (Alderson, 2000), Alderson and Lukmani's (1989) eight- subskill hierarchy, Heaton's (1991), pp. 14-subskill taxonomy, and Hughes (2003), pp. 23-subskill checklist. Appendix A displays the five subskill lists with similar subskills grouped together.

Cognitive levels of reading subskills

Reading is essentially a dimension of thinking (Pearson & Raphael, 1990). The "simple view reading" theory (Hoover & Gough, 1990) also argued that most reading comprehension skills were essentially cognitive or thinking skills. Thinking activities are hierarchical. The Taxonomy of Educational Objectives (Bloom et al., 1956) is a broadly recognized classification assessing thinking skills in education. Bloom's Taxonomy initially comprised six levels of thinking skills which are Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation, ranging from simple to complex or from concrete to abstract mental processing abilities. Anderson and Krathwohl (2001) proposed a revised version of Bloom's taxonomy, which comprises remembering, understanding, applying, analyzing, evaluating, and creating. The first three levels are categorized as lower-order thinking skills while the last three levels are classified as higher-order thinking Skills (Moore & Stanley, 2013). Table 1 contains the definitions of the six thinking skills.

Thinking skills are general cognitive processes that exist in all kinds of thinking activities, including reading. As specific representations of thinking skills in reading, reading subskills can also be classified into different cognitive levels. Luebke and Lorié (2013) classified reading comprehension questions into four categories: (1) Recognition, (2) Understanding and Analysis, (3) Inference, and (4) Application. Those categories are intended to represent a hierarchy of reading subskills, with the latter categories representing higher cognitive levels.

Closely related to cognitive levels in reading is the notion about local or global comprehension (Urquhart & Weir, 2014; Weir & Khalifa, 2008). In many cases, thinking skills involving complex thinking activities deal with understanding at global or text levels and thinking skills involving simple thinking activities deal with understanding at local or sentence levels. However, global comprehension is not necessarily at higher cognitive levels

Table 1. Definitions of thinking skills in revised Bloom's taxonomy.

		Bloom's Definition
Lower-Order Thinking Skills	Remembering	Exhibiting memory of previously learned material by recalling facts, terms, basic concepts, and answers.
	Understanding	Demonstrating understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating main ideas.
	Applying	Solving problems to new situations by applying acquired knowledge, facts, techniques and rules in a different way.
Higher-Order Thinking Skills	Analyzing	Examining and breaking information into parts by identifying motives or causes. Making inferences and finding evidence to support generalizations.
	Evaluating	Presenting and defending opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.
	Creating	Compiling information together in a different way by combining elements in a new pattern or proposing alternative solutions.

The table is adapted from *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives* (Anderson & Krathwohl, 2001).

and local comprehension is not necessarily at lower cognitive levels. The reason may lie in that cognitive levels emphasize characteristics of thinking while local or global comprehension focuses on the language structure under analysis. The notion about global and local comprehension is relative and can be described as a continuum.

Reading process models

The hierarchical characteristics of reading subskills are also reflected in the theories on reading processes. Since the 1960s, scholars have proposed a series of reading process models. The most influential ones include the information processing model (Gough, 1972), the psycholinguistic model (Goodman, 1967; Smith, 1977), the interactive model (D. E. Rumelhart, 1977; D. Rumelhart, 1980), and the construction-integration (CI) model (Kintsch, 1998, 1988).

The information processing model (Gough, 1972), also known as the bottom-up model, proposed a progressive comprehension process from decoding to meaning construction. The psycholinguistic model (Goodman, 1967; Smith, 1977) is exactly the opposite of the bottom-up model. Larger units affect the way smaller units are perceived. In contrast with the one-way linear transmission of information proposed by the above two models, the interactive reading model (D. E. Rumelhart, 1977) posits that the reading process is a dialogue process between the reader and the text. Closely related to the interaction model is the schema theory (D. Rumelhart, 1980) which claimed that the most effective reading comprehension mode should be a combination of "bottom-up" and "top-down" modes, and the two information processing modes should co-occur and affect each other at all levels.

The above reading process models can also be reflected in the construction-integration (CI) model (Kintsch, 1998, 1988). According to Kintsch (2005), the construction-integration model describes the interplay between top-down and bottom-up processes in comprehension. The construction-integration model divides the process of text reading comprehension into two stages: construction and integration. The construction stage refers to the formation of the textbase, which is a literal and exact structural, as well as semantic representation of the text. The integration stage refines the textbase and forms the situational model, which is a representation of the text that is achieved through the integration of textbase with prior knowledge through elaboration and inference processes. Although the CI model was initially developed for L1 reading, Nassaji's studies (Nassaji, 2002, 2007) confirmed its applicability in L2 reading. L2 readers and L1 readers share similar text processing skills but L2 readers may lack efficiency of processing, especially in construction.

The interplay between bottom-up and top-down patterns was also reflected in studies on the compensatory nature of the reading process. Stanovich (1980) first proposed the notion of compensatory processing in reading. Bernhardt (2005) and McNeil (2012) discussed the idea in more detail. The former found that the L1 competence (reading strategies) and the L2 competence (L2 linguistic knowledge) may compensate each other collaboratively and interactively in the L2 reading process. McNeil (2012) believes that the L2 reading process is also influenced by the compensatory relations between the L2 background knowledge which involves the integration of the textbase with the reader's knowledge and the L2 strategic knowledge which encompasses the conscious cognitive and metacognitive mental actions readers take to plan, repair, evaluate, and monitor comprehension processes.

Processing larger units or processing smaller units cannot be regarded as a subskill because such processing is too general to demonstrate the distinctive characteristics of reading. Although the notions of textbase, background knowledge, and strategic knowledge may somewhat reflect features of reading process, they are still not fine-grained and thus cannot provide detailed diagnostic information for teaching and learning. Fine-grained reading subskills are the focus of many quantitative studies.

Quantitative studies on reading subskill associations

The methods adopted by conventional quantitative studies on reading subskill associations mainly include factor analysis, multiple regression, multitrait-multimethod correlational analysis, and structural equation modeling. Although studies conducted with factor analysis (Liao et al., 2016; Pierce et al., 2010) and multiple regression (Davey, 1988) could identify clusters of items belonging to the same reading subskill, those methods could not reveal how reading subskills were related and what the directions of such associations were. Moreover, although studies conducted with multitrait–multimethod correlational analysis (MTMM; Dickinson & Adelson, 2016; Shermis & Long, 2009) and structural equation modeling (Muijselaar et al., 2017; Van Steensel et al., 2013) might, to some extent, detect the association among reading subskills, those methods could not deal with multiple subskills represented by a single item, which might not cater to the nature of reading skills that are highly integrative and closely related.

It is also not appropriate to use unidimensional IRT models or multidimensional IRT models to explore reading subskill associations because both methods focus on continuous latent traits calibrated on the interval scale of ability, which meets the need to compare ability among students. Cognitive diagnosis, in contrast, focuses on discrete attributes and thus can be adopted to detect the mastery/nonmastery of subskills (Hong et al., 2015).

Cognitive diagnosis of reading subskills

The purpose of cognitive diagnostic assessment is to measure or evaluate an individual's specific knowledge structure and patterns of attribute processing (Gierl et al., 2000). Cognitive diagnosis obtains the observable response pattern of the examinee through assessment to estimate the unobservable knowledge state of the examinee. Cognitive diagnosis, which is sensitive to estimating finer-grained latent variables, is a substantive assessment method designed to explore the cognitive process. It can find the differences in the internal cognitive process or knowledge structure among examinees so as to provide individualized teaching and personalized interventions. Moreover, it is a method that can analyze multiple subskills represented by a single item. Based on psychometric modelling and assumptions about cognitive processes, cognitive diagnosis models (CDMs) were developed to realize the measurement of the internal cognitive process of examinees and provide cognitive diagnostic information.

Since a reading item may represent multiple interrelated reading subskills, cognitive diagnosis, which is sensitive to estimating finer-grained latent variables, is an ideal method to explore reading subskills. Many attempts have been made to study reading subskills using various CDMs. Von Davier (2008), Lee and Sawaki (2009), Jang (2009), and Wang and Gierl (2011), and Kim (2015) conducted cognitive diagnosis analyses of reading comprehension subskills using the general diagnostic model (Von Davier, 2005), reduced reparameterized unified model (R-RUM; Hartz, 2002; Hartz & Roussos, 2008), and attribute hierarchy method (Leighton et al., 2004). However, implicitly or explicitly, the CDMs involved in these analyses are considered reduced models, in that they are special cases or constrained versions of a more general CDM. As such, when a general CDM is appropriate, reduced models could not fully represent how the different subskills interact to produce a correct response. Adopting to the current context, this can be referred to as item-level subskill interaction (de la Torre et al., 2018). For example, the R-RUM is considered a reduced model, where the contribution of a subskill is deemed independent of the contributions of the other subskills (de la Torre, 2011). In other words, this model assumes that the probability of correctly responding to an item can be fully described without including the effects arising from the possible interactions between the subskills. This may not cater to the features of reading subskills. Recently, some scholars noted that the item-level interaction(s) of the reading subskills might not be known a priori, thus, cannot be assumed. In such a situation, a general or saturated CDM should be adopted because it can estimate all the possible item-level subskill interaction effects. Thus, it has sufficient flexibility to accommodate various interactions between the subskills (Li et al., 2016). Many studies (e.g., Chen & Chen, 2013, 2016a, 2016b; Ravand & Robitzsch, 2018) have adopted the saturated generalized deterministic inputs, noisy "and" gate (G-DINA) model to diagnose reading subskills and have confirmed its feasibility. As for the CDM research on reading subskill associations, Chen and Chen (2016a) proposed a network of subskill associations by simply counting the occurrence of subskill coexistence based on latent classes.

Research questions

Previous studies on the associations among specific reading subskills were limited in two ways. First, the traditional quantitative methods mentioned in the literature review could not fully detect subskill relationships (e.g., regression, factor analysis, MTMM) or do not have appropriate assumptions (e.g., unidimensional and multidimensional IRT). Second, most of the CDM research focused on the method or accuracy of diagnosis because some studies mainly paid attention to comparison and selection among different models for diagnosis (e.g., Li et al., 2016; Ravand & Robitzsch, 2018) while others attempted to find out whether CDMs could diagnose language skills accurately (e.g., Jang, 2009; Kim, 2015; Lee & Sawaki, 2009; Von Davier, 2008; Wang & Gierl, 2011). However, all the CDM studies on reading subskills have paved the way for further research on reading theories. To explore the reading subskill associations from a CDM perspective and promote theories on the reading construct, this study proposed the following research questions:

(1) To what extent can reading subskill associations be disclosed using the CDM approach?
(2) How are these reading subskills related to each other?

Methods

Data description

We adopted one reading comprehension subtest from Band 8 of Test for English Majors (TEM8). The Test for English Majors (TEM) is an English proficiency test program in China designed to assess the English language proficiency of undergraduate English majors. It is used to determine whether examinees have met the learning requirements specified in the teaching syllabus and provide teachers with feedback on their teaching effectiveness and students on their strengths and weaknesses in English learning. Although TEM cannot be regarded as diagnostic, the test data can also be used for CDM analysis since retrofitting attempts can also retrieve useful diagnostic information and serve as necessary steps to advance CDM research (Lee & Sawaki, 2009). The test has two levels: TEM4, which is required to be taken by sophomore English major students, and TEM8, which is administered to senior English major students. TEM4 and TEM8 are separate tests administered annually to students at the two stages of university study. They are roughly at the B2 and C1 levels of Common European Framework of Reference for Languages (CEFR), respectively (Liu & Wu, 2019; Yang & Liu, 2019). Since the CDM research adopting C-level subjects is quite limited and learners of higher levels may possess more reading subskills, we chose TEM8 candidates as the participants of this study.

TEM8 candidates are senior students of undergraduate English major in China. They are required to understand the editorials and book reviews in British and American newspapers and magazines and appreciate biographies and literary works suitable to advanced learners in China (around C1). The reading subskills required for TEM8 candidates include understanding the main ideas, distinguishing facts and details, understanding explicit information and inferring implicit information, making evaluation and deduction, and analyzing discourse structure, language characteristics, and rhetoric, as stated by *Teaching syllabus for English majors* (English Major Division of National Foreign Languages Advisory Board, 2000). The design of TEM8 reading is strictly based on the above requirements and the same subskills were reiterated in *Test syllabus of TEM8* (Revision Team of TEM8 Syllabus, 2004). According to the syllabi on which TEM8 reading is constructed, reading subskills in TEM8 can be identified and therefore diagnostic information at subskill levels can be obtained.

The TEM8 reading comprehension subtest contains four articles and 20 dichotomously scored items. The four articles are a review about recent trends of university development, an excerpt of an American novel, an excerpt of an autobiography, and a report about the history of museums respectively. The length of each article is about 600 words and the average of Flesch Kincaid Grade Level Readability (Kincaid et al., 1975) is 12.38 which is close to college level for native English speakers. The 20 items are all multiple-choice items with four options.

The TEM Examination Board provided the authors with a dataset including 1203 test takers randomly sampled from all the 123,730 TEM8 candidates taking the same Test for English Majors nationwide. Random sampling was carried out by setting a selection proportion at about 1% in SPSS case selection function. The sample was larger than those adopted in some recent CDM studies on reading (e.g., Du & Ma, 2021; Fan & Yan, 2020).

G-DINA model

The G-DINA model (de la Torre, 2011) is a saturated CDM and its formula can be written as follows:

$$P\left(\alpha_{ij}^{*}\right) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{ik} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{ik}\alpha_{ik'} + \cdots + \delta_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{ik}.$$

In this formula, $P(\alpha_{ij}^{*})$ represents the probability of a correct response of student i with subskill (or attribute) pattern α_{ij}^{*} on item j that requires K_j^* subskills, δ_{j0} represents the baseline probability of a correct response, δ_{jk} represents the main effect of mastering subskill k, $\delta_{jkk'}$ the interaction effect of simultaneously mastering subskills k and k', and $\delta_{j12\cdots K_j^*}$ the interaction effect of mastering all the required subskills for the item. The formula contains all the possible effects (i.e., baseline, all main effects, and all possible interaction effects). Thus, the formula represents the saturated model. With appropriate constraints, different reduced or simpler CDMs can be derived from the saturated model. The saturated model could cover both compensatory skill relationships (a high level of competence on one skill/subskills can compensate for a low level of competence on another in answering the test items) and non-compensatory skill relationships (no compensation exists between skills/subskills; Javidanmehr & Anani Sarab, 2019) and estimate hierarchically structured attributes (Akbay & de la Torre, 2020) because the saturated model includes all possible interactions among attributes. Previous studies (Chen & Chen, 2013, 2016a, 2016b; Ravand & Robitzsch, 2018) have shown that the G-DINA model can provide adequate fit to reading assessments. Hence, the G-DINA model is adopted in this study. The CDM analysis was carried out through the G-DINA analysis program based on Ox program developed by de la Torre (2011). The steps in a CDM analysis involves subskill definition, Q-matrix construction, model fit, attribute mastery prevalence, and latent class size estimations, which will be elaborated about in the section of CDM analysis.

Model validation

Subskill definition

Since TEM8 reading subskills were not defined theoretically, we adopted the 19 micro-skills of reading put forward by Munby (Alderson, 2000) and the 20-subskill list for careful reading formulated by Hughes (2003) as references to analyze the TEM8 reading subskills. We chose the two classifications because they are highly granular and discussed in the language testing handbook compiled by the TEM expert team (Zou, 2012). We examined the two lists of reading subskills and found that the subskills in both lists could be merged into 11 subskill categories, which were taken as the initial reading subskills for experts to conduct content analysis on the TEM8 test items used in this study in order to determine the reading attributes of each item. Table 2 shows the reading subskills, their sources, their codes, and their definitions.

In Table 2, it can also be found that the two scholarly sources overlap with TEM8 reading requirements for most of the subskills, which demonstrates that the two scholarly classifications may constitute the theoretical framework of TEM8 reading subskills.

Table 2. Definition of reading subskills.

Codename	Subskill	Definition	Source
Recognize	Recognizing words and phrases	Recognizing the words and phrases affecting reading comprehension	Munby: Recognizing the script of a language.
Grammar	Understanding grammatical structures	Understanding the sentence structures affecting reading comprehension	Hughes: Interpreting complex sentences. Munby: Understanding relations within the sentence. TEM8 requirements: Analyzing language characteristics.
Meaning	Making inference of unfamiliar words	Inferring meanings of unfamiliar words according to context and word formation	Hughes: Inferring the meaning of an unknown word from context. Munby: Deducing the meaning and use of unfamiliar lexical items.
Explicit	Retrieving explicitly stated information	Locating and retrieving explicit fact information at single location	Munby: Understanding explicitly stated information. TEM8 requirements: Understanding explicit information.
Interpret	Information interpretation	Explaining and paraphrasing conceptual meanings at single location	Munby: Understanding conceptual meaning. TEM8 requirements: Making deduction.
Multi	Interpreting multiple details	Interpreting information from multiple locations, usually involving distinguishing or comparison	Hughes: Extracting relevant points from a text selectively. Munby: Subskills involving distinguishing. TEM8 requirements: Distinguishing facts and details.
Summary	Identifying main ideas	Identifying the main ideas of paragraphs or the whole text according to explicit information	Hughes: Interpreting topic sentences, Identifying explicitly stated main ideas. Munby: Extracting salient details to summarize (the text, an idea), Identifying the main point or important information in discourse. TEM8 requirements: Understanding the main ideas.
Cohesion	Understanding textual cohesive relationships	Understanding and recognizing cohesive and referential relationships	Hughes: Identifying pronominal reference, Identifying discourse markers Munby: Recognizing indicators in discourse, Understanding relations between parts of text through lexical cohesion devices, Understanding cohesion between parts of a text through grammatical cohesion devices.
Infer	Making inference on content	Making inference on content according to background knowledge	Hughes: Making pragmatic inferences. Munby: Understanding information when not explicitly stated. TEM8 requirements: Inferring implicit information.
Intent	Inferring author's purpose, attitude, mood, and emotion	Inferring author's purpose, attitude, mood, and emotion based on the whole text	Hughes: Recognizing writer's intention, Recognizing the attitudes and emotions of the writer.
Style	Evaluating text types	Evaluating text types according to the rhetorical structure of the passage	Hughes: Identifying what kind of text is involved. TEM8 requirements: Analyzing discourse structure and rhetoric, Making evaluation.

The subskills from Hughes' list were quoted from *Testing for language teachers* (Hughes, 2003, p. 139).
The subskills from Munby' list were quoted from *Assessing reading* (Alderson, 2000, pp. 10–11).
The subskills required by TEM8 were adapted from *English teaching syllabus for English majors* (English Major Division of National Foreign Languages Advisory Board, 2000).

Q-matrix construction

We invited five teachers (one associate professor and four lecturers) and two graduate students who had degrees on applied linguistics and experiences in teaching English majors to code the subskills independently for the reading items. We conducted an interrater agreement survey on each subskill coding of each item based on the coding results of those teachers and graduate students. If more than half (i.e., four or more) of the above seven raters agree with a certain subskill coding of an item, then the subskill coded is recognized as a reading comprehension subskill represented by that particular item. If less than half agree, then the subskill coded is regarded as invalid. Based on the coding results of the seven experts, we obtained a coding matrix with an average agreement percentage at 86% (6 out of 7). According to the agreed coding matrix, the TEM8 items represent a total of eight reading subskills. Three reading subskills defined earlier (Recognize, Grammar, and Cohesion) were dropped because few experts coded them based on the TEM8 reading subtest in study. The probable reason might be that Grammar was regarded as the basics of other subskills, and Recognize and Cohesion were not explicitly stated in TEM8 reading requirements. Item examples for the eight subskills are provided in Appendix B.

Based on the definitions of the thinking skills of the revised version of Bloom's taxonomy (Anderson & Krathwohl, 2001) shown in Table 1 and the descriptions of Luebke and Lorié's (2013) reading question categories, a continuum of cognitive level was established for the eight reading subskills. Both Explicit and Interpret are concerned with explicit information at single location but the latter involves explaining and paraphrasing conceptual meanings. Multi is similar to Interpret but involves information from multiple locations. Meaning is partially concerned with vocabulary knowledge and partially concerned with inference based on local context (Huckin & Bloch, 1993). Summary here only refers to identifying main ideas, which is lower than inference according to the level of cognitive processing

Table 3. Cognitive level continuum of reading subskills in TEM8.

Reading subskill	Codename	Cognitive level
Retrieving explicitly stated information	Explicit	low
Information interpretation	Interpret	
Interpreting multiple details	Multi	
Making inference of unfamiliar words	Meaning	
Identifying main ideas	Summary	
Making inference on content	Infer	
Inferring author's purpose, attitude, mood, and emotion	Intent	
Evaluating text types	Style	High

Table 4. Q-matrix for TEM8.

	Meaning	Explicit	Interpret	Multi	Summary	Infer	Intent	Style
1	0	0	0	0	1	0	1	0
2	0	0	1	1	0	0	0	0
3	0	1	0	1	0	1	0	0
4	0	0	0	1	1	0	0	0
5	0	0	0	0	0	0	1	1
6	0	0	0	1	0	1	0	0
7	0	1	0	1	0	0	0	0
8	0	0	1	1	0	1	0	0
9	0	0	0	1	0	1	0	0
10	1	0	0	1	0	0	0	0
11	0	1	0	1	0	0	0	0
12	0	1	1	0	0	0	0	0
13	0	1	1	0	0	0	0	0
14	0	0	0	0	1	0	1	1
15	1	0	0	0	1	1	0	0
16	0	1	1	0	0	0	0	0
17	1	0	1	0	0	0	0	0
18	0	1	1	0	0	0	0	0
19	0	0	1	1	0	0	0	0
20	0	0	0	0	1	0	0	0
Summ	3	7	8	10	5	5	3	2

(Grabe, 2009). Infer is about making inference based on background knowledge. Intent involves synthesizing. Style is concerned with evaluation of text types which are essentially the rhetorical structures or patterns in texts (Paltridge, 1996). We did not assign definite cognitive levels to reading subskills which were just relatively higher or lower at cognitive levels. Table 3 shows the cognitive level continuum of the eight reading subskills in TEM8.

The refining process of the coding matrix was carried out through a series of estimations with the G-DINA analysis program based on Ox program developed by de la Torre (2011). To detect the item-level misspecifications, we adopted two statistics: *zr*, which is the standardized residual between the observed and predicted Fisher transformed correlations between an item pair; and *zl*, which is the standardized residual between the observed and predicted log-odds ratios of an item pair. By averaging all the *zr*s related to item *j*, we obtained $sr(j)$. Similarly, we obtained $sl(j)$ by averaging all the *zl*s related to item *j*. The item with the maximum $sr(j)$ or $sl(j)$ was considered to be most likely mis-specified. According to the consensus of the above experts, modification was made to the coding of the most likely mis-specified item which could be clearly identified by the program based on Ox in each estimation. Chen (2017) elaborated on the details of the process. The process continued until we achieved test-level absolute model fit (at $p = .05$ level).

The seven experts then gathered together to discuss the three subskills (Meaning, Intent, and Style) with the smallest coding frequency (two) and agreed that one more coding could be made for Meaning and Intent. Therefore, all subskills except Style were coded at least three times. Although Style was coded only twice, we still decided to keep it. The reasons are that the items assessing Style can be found in most reading tests though the proportion of such items is always very low and deleting the subskill may lead to restriction of construct. We ran the CDM analysis again and found the absolute model fit was further improved (at $p = .10$ level). The final Q-matrix for the TEM8 reading items is presented in Table 4.

Model fit of cognitive diagnosis

Still based on the *zr* and *zl* statistics, the absolute fit of G-DINA model to the data was evaluated at the test level. Under the null hypothesis that the model fits the data, the residuals are hypothesized to be equal to 0. According to Chen et al. (2013), the model fit can be deemed satisfactory if the maximum *zr* and *zl* statistics are less than the Bonferroni adjusted critical z-score *zc* at a particular significance level. In this study, the significance level was finally set at $p = .10$, which is equivalent to the Bonferroni adjusted $zc = 3.47$, and the maximum *zr* and *zl* statistics were 3.02 and 2.95, respectively. Therefore, it can be concluded that the Q-matrix given in Table 4 in conjunction with the G-DINA model can adequately fit the TEM8 sample data. We also examined the fit of the G-DINA model against five reduced CDMs (i.e., DINA model, DINO model, RRUM, A-CDM and LLM) using −2LL, AIC, and BIC, and found that the fit indices under the G-DINA model were smaller than almost all the fit indices under the reduced models, which indicates a better fit under the G-DINA model (Li et al., 2016; Ravand & Robitzsch, 2018). The only exception is that the BIC under the G-DINA model was slightly larger than that under the LLM model which is also a compensatory model. The better fit of the G-DINA model may also provide evidence that the assumptions of the saturated G-DINA model cater to the features of reading subskills which are interactive and compensatory, as was elaborated about in the literature review. As such, results reported in this study were based on the G-DINA model.

Establishing the reading subskill associations

An important output of the CDM analysis is the table containing the relative sizes of the different latent classes estimated from the sample. Each latent class represents a subskill profile, and is characterized by a vector of 1 and 0 corresponding to the *K* subskills, where 0 and 1 indicate the nonmastery and mastery of the subskills, respectively. The frequency of mastering a subskill or a pair of subskills could be obtained by summing up the frequencies of all latent classes containing the subskill or the pair of subskills. The frequencies, or test-takers' posterior latent class membership, were analyzed in the chi-square test of independence and odds ratio estimation because chi-square test and odds ratio can be used to analyze binary data (mastery or nonmastery) according to the occurrence of categories (latent classes containing relevant subskills). In this study, we used the chi-square test of independence to examine whether an association exists between two subskills. Specifically, the chi-square test of independence was conducted for all possible pairs of subskills by employing the Crosstabs function in SPSS. For significantly related subskills, we also computed the odds ratio (θ) to quantify the strength of association between the two subskills (Agresti, 2002; Gilbert, 1993) by estimating the Risk, which is an optional statistic under the Crosstabs function. Assuming subskills A and B are related, the process of calculating the odds ratio between subskill A and subskill B can be illustrated in a two-by-two frequency table.

a	b
c	d

a = Frequency of mastering both A and B
b = Frequency of mastering B not A
c = Frequency of mastering A not B
d = Frequency of mastering neither A nor B

Based on the frequencies in the four slots, the odds ratio (θ) can be calculated in the following equation.

$$\theta = \frac{a/c}{b/d}$$

An odds ratio greater than 1 in this context indicates that, compared to those who did not master subskill A, students who did master subskill A are more likely to master subskill B; an odds ratio less than 1 indicates that students who have mastered subskill A are less likely to master subskill B. In CDM research, odds ratio was also used to detect differential item functioning (DIF) and differential attribute functioning (DAF; Li, 2008).

Another important output of the CDM analysis is the attribute mastery prevalence, which can be defined as the overall proportion of participants who master a certain attribute. We established the stochastic direction of each significantly correlated associations on the basis of attribute mastery prevalence. In establishing the stochastic direction, the attribute with higher mastery prevalence is regarded as the likely antecedent of the attribute with lower mastery prevalence. This is different from a deterministic association,

Table 5. Latent classes with the largest sizes for the TEM8 sample.

Pattern	Latent classes	Size(%)	Frequency
1	11,111,111	19.87%	239
2	01111111	5.12%	62
3	10,001,001	3.94%	47
4	00001110	3.88%	47
5	00000010	2.91%	35
6	11,011,010	2.81%	34
7	00000001	2.26%	27
8	00000000	2.26%	27
9	10,101,110	2.25%	27
10	10,000,010	2.16%	26
11	10,000,111	2.10%	25
12	10,111,010	1.96%	24
13	10,101,010	1.92%	23
14	10,010,101	1.61%	19
15	01101010	1.45%	17
16	00101010	1.45%	17
17	01011010	1.43%	17
18	11,011,000	1.38%	17
19	10,011,101	1.38%	17
20	00000110	1.34%	16
21	00001010	1.29%	16
22	00010110	1.26%	15
23	00010101	1.18%	14
24	11,010,110	1.09%	13
25	10,011,001	1.09%	13
26	01001100	1.07%	13
27	01001000	1.07%	13
28	00011110	1.06%	13
29	00000101	1.01%	12

The order of the subskills in the latent classes is Meaning, Explicit, Interpret, Multi, Summary, Infer, Intent, and Style from left to right.
Only the 29 latent classes with a size of 1% or larger are displayed.

which assumes that mastery of an easier reading subskill is strictly necessary for one to master a more difficult subskill.

Based on the above-mentioned output, a network of reading subskill associations can be established. According to the results of the chi-square test, the correlated reading subskills can be linked by arrow lines whose directions are determined by the stochastic directions set by the attribute mastery prevalence. The arrow lines are marked with the subskill pairs' odds ratios which stand for the strengths of the relevant associations. The attribute mastery prevalence of each subskill can also be displayed to show each subskill's degree of dominance in the process of subskill mastery.

Results

Latent class size estimates

One of the advantages of CDM is that it can estimate the latent class sizes from the sample. Latent classes reflect the various patterns of subskill combination in student cognition. The larger the size of a latent class, the more students possess the associated subskill combination.

Table 5 displays the latent class sizes for the TEM8 sample. With eight attributes, there are theoretically 256 latent classes. However, Table 5 only displays the 29 latent classes with a size of 1% or larger – these latent classes account for 73.6% of the cases.

Table 5 shows that 19.87% of the students have mastered all the eight reading subskills, and slightly more than 2% did not master any of the subskills. From Table 5, it can be observed that the reading subskill patterns of the TEM8 sample were quite diversified, which may be due to the wide variety of extracurricular reading practices junior and senior students engaged in. Universities in China usually do not offer intensive or extensive reading courses for junior and senior English majors who, in turn, practice reading after class or in various academic courses. Based on the latent class size estimates, we retrieved information about the frequencies of subskill mastery, which were used for conducting the chi-square test and odds ratio estimation.

Strength of subskill associations

Based on the CDM output, the associations among the reading subskills were explored by computing the chi-square test and odds ratio. The latent classes whose sizes are smaller than 1% of the total sample were also included in the computation. Table 6 displays the results of chi-square tests of independence for all possible pairwise associations and the strength of the significant associations.

Since multiple comparisons were conducted simultaneously, the Bonferroni correction was applied to the chi-square and odds ratio analyses. Specifically, the original *p*-values were multiplied by 28, which is the number of comparisons. Thus, the symbol ** indicates that the *p*-value was significant at alpha = .05 after the Bonferroni correction. Among the 28 associations, 21 were found significantly correlated after Bonferroni correction. It can be seen that one subskill is usually much more global than the other in a pair of significantly associated subskills (e.g., Intent, Summary, and Multi at passage or paragraph levels are significantly associated with Explicit, Interpret, and Meaning at sentence or word levels).

Table 6. Chi-square tests and odds ratios of the reading subskills.

	Explicit	Interpret	Multi	Summary	Infer	Intent	Style
Meaning	$\chi^2 = 25.67$* *$\theta = 1.88$**	$\chi^2 = 26.27$* *$\theta = 1.84$**	$\chi^2 = 45.84$* *$\theta = 2.23$**	$\chi^2 = 88.09$* *$\theta = 3.39$**	$\chi^2 = 33.45$* *$\theta = 1.98$**	$\chi^2 = 90.30$* *$\theta = 3.14$**	$\chi^2 = 34.47$* *$\theta = 2.01$**
Explicit		$\chi^2 = 0.58$	$\chi^2 = 31.60$* *$\theta = 1.98$**	$\chi^2 = 55.87$* *$\theta = 2.47$**	$\chi^2 = 1.84$	$\chi^2 = 83.11$* *$\theta = 2.97$**	$\chi^2 = 0.46$
Interpret			$\chi^2 = 6.63$* *$\theta =1.94$**	$\chi^2 = 42.11$* *$\theta =2.44$**	$\chi^2 = 1.29$	$\chi^2 = 82.90$* *$\theta =2.99$**	$\chi^2 = 0.21$
Multi				$\chi^2 = 27.00$* *$\theta = 2.04$**	$\chi^2 = 26.91$* *$\theta = 1.85$**	$\chi^2 = 50.38$* *$\theta = 2.39$**	$\chi^2 = 11.28$* *$\theta = 1.62$**
Summary					$\chi^2 = 35.23$* *$\theta = 2.26$**	$\chi^2 = 0.37$	$\chi^2 = 42.40$* *$\theta = 2.17$**
Infer						$\chi^2 = 60.67$* *$\theta = 2.54$**	$\chi^2 = 1.78$
Intent							$\chi^2 = 58.83$* *$\theta = 2.51$**

**indicates $p < 0.05$ after Bonferroni correction; χ^2 = Pearson chi-square statistic; θ = odds ratio.

Although the chi-square test of independence can determine whether there is a significant correlation between two categorical variables, it cannot estimate the strength of the association. Therefore, the odds ratio was computed for significantly correlated associations. Although the odds ratio value can range from 0 to infinity (Gilbert, 1993), the odds ratios for significantly correlated associations in Table 6 are all larger than 1. Thus, the odds ratios for the reading subskills of the sample can be interpreted as follows: compared to those who have not mastered the first subskill, students who have mastered the first subskill are more likely also to have mastered the second subskill. For example, if we take Meaning as the first subskill and Summary as the second subskill, students who have mastered Meaning were $\theta = 3.39$ times more likely to have mastered Summary compared to students who have not mastered Meaning. Table 6 displays a diagonal matrix. The rows represent the first subskills and columns represent the second subskills instead of a full matrix because the reverse between rows and columns makes no difference to the odds ratio values.

Stochastic directions of associations

To determine the stochastic direction of the association, we examined the overall subskill prevalence. Attribute mastery prevalence, an output of CDM analysis, defines the overall proportion of subjects who master a certain attribute – an attribute that has a higher attribute mastery prevalence is easier to master; conversely, an attribute with lower mastery prevalence is more difficult to master. Intent was the easiest to master, with a 77% prevalence, followed by Summary (71%), Multi (63%), Style (59%), Infer (58%), Interpret (56%),HT Explicit (56%), and Meaning (51%).

According to attribute mastery prevalence, stochastic directions between subskills could be established because easier reading subskills are likely the antecedent of more difficult subskills. It should be noted the term "likely antecedent" was used rather than "prerequisite" because we do not assume that mastery of an easier reading subskill is strictly necessary for one to master a more difficult subskill. The phenomenon that subskills at different difficulty levels are associated at least warrants the existence of dependency or co-occurrence between an easy subskill and a difficult one.

Table 6 and elaborations on attribute mastery prevalence were summarized diagrammatically in Figure 1, which shows both the stochastic directions among the subskills and strength of the associations.

Figure 1 demonstrates that the CDM approach can establish a comprehensive network of subskill associations. The subskills with arrow lines directing to many other subskills can be regarded as easier. Intent and Summary each possessed more such arrow lines than other subskills each possessed, which demonstrated that those two subskills could be the easiest subskills for the sample. We found discrepancies between difficulty of some subskills and their cognitive levels. The subskills on the left side of Figure 1 were roughly easier than those on the right side, whereas the former tended to be more cognitively challenging than the latter.

Among all the 21 reading subskill associations for the sample, eight subskill associations (Summary→Infer, Summary→Style, Intent→Style, Multi→Meaning, Multi→Infer, Multi→Style, Explicit→Meaning, and Interpret→Meaning) were in the pattern in which the preceding subskill is lower in cognitive levels than the succeeding subskill; and 13 subskill associations (Intent→Explicit, Intent→Interpret, Intent→Meaning, Intent→Multi,

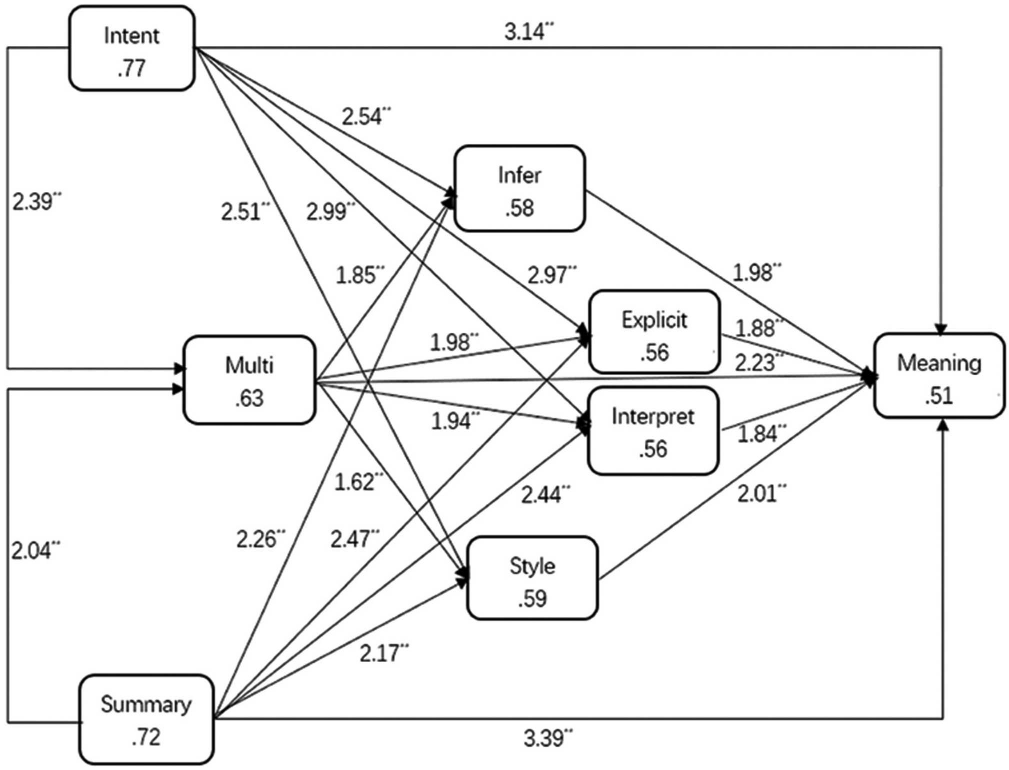

Figure 1. Reading subskill associations for the TEM8 sample. Number in the subskill box indicates the mastery prevalence or difficulty of the subskill.Number on the arrow line indicates the odds ratio (strength) of the subskill association.→represents the stochastic direction of significant subskill association.**indicates $p < .05$ after Bonferroni correction.

Intent→Infer, Summary→Explicit, Summary→Interpret, Summary→Meaning, Summary→Multi, Multi→Explicit, Multi→Interpret, Infer→Meaning, and Style→Meaning) were in the pattern in which the preceding subskill is higher in cognitive levels than the succeeding subskill. The two patterns co-occurred in the subskill associations for the TEM8 sample but the pattern in which the preceding subskill is higher in cognitive levels than the succeeding subskill was more dominant.

Discussion

Disclosing reading subskill associations through CDM

The CDM analysis provides the output of latent class size estimates and attribute mastery prevalence. Based on latent class size estimates, the chi-square test of independence can help discover the significantly correlated subskill associations, and odds ratio estimation can help determine the strengths of those associations. Based on attribute mastery prevalence, we can also establish the stochastic directions between reading subskills. Therefore, with the output produced from the CDM analysis, we can chart a network of reading subskills displaying significant links, link strengths, and link directions, demonstrating that the CDM approach has the capacity to investigate the reading subskill relationship in an objective and comprehensive way. In this study, 21 out of all the 28 subskill associations were found statistically significant and stochastically ordered subskill relationships based on cognitive levels were detected, which is in line with the G-DINA model assumptions that attribute relationships are interactive and compensatory.

Reading subskill association patterns

The 21 significant associations also demonstrated the co-occurrence of the two opposite patterns based on cognitive levels. Although the two patterns of subskill associations are derived from stochastic directions between subskills, the existence of the two patterns can hardly be denied since many associations belong to either pattern. The co-occurrence of the two opposite patterns may partially reflect the construction-integration model (Kintsch, 1998, 1988), which demonstrates that reading process is the interplay between the construction process for decoding and the integration process for higher-order thinking. However, the subskill pair pattern in which the preceding subskill is higher in cognitive levels than the succeeding subskill was possibly more dominant in the TEM8 sample. That pattern seems to be more suitable to the sample students because L2 learners may be inclined to use textual clues to compensate for slower lexical access due to inefficient decoding (Stanovich, 1980) even though they are advanced English learners in China.

Although the reading subskill associations established through the CDM approach (based on reading subskill mastery statistics) may be interpreted as a reflection of the reading subskill learning process, they could be to some extent explained by some major reading process models. The probable reason is that mastering general skill of reading comprehension and the component skills is largely based on readers' reading experience (Perfetti et al., 2005).

In addition to the stochastically ordered subskill relationships based on cognitive levels, the research on reading subskill association also revealed a reading subskill hierarchy which is determined by the degree of overall comprehension. The study discovered that the participants tended to master global comprehension subskills more easily than master comprehension subskills focusing on more specific information. It seems that difficult words and sentences were the real obstacles in L2 reading even for advanced learners in China, which echoes with research on TEM reading (Tian, 2008; Zou et al., 2002) that the subskills concerning the understanding of words and sentences may lead to lower reading scores of senior English major students in China. This finding also reflects Nassaji's discovery (Nassaji, 2002, 2007) that L2 readers lack efficiency in construction because they spend more time processing smaller linguistic units. The obvious exception is that the more global subskill Style is slightly more difficult than Multi which is less global. Probably fewer items coded for Style may make the estimation less accurate.

In this study, one subskill is usually much more global than the other in a pair of significantly associated subskills and the more global subskill is usually the preceding subskill. According to Urquhart and Weir (2014), larger units concerning background knowledge may compensate for local grammatical processing. It seems that the compensation between the comprehension at a more global level and the comprehension at a less global level may contribute to significant association especially when competences needed at the two levels are quite different. The lack of association between Infer and Style is probably attributable to the weak compensation between them because both subskills are global comprehension subskills and they are at the same level of difficulty, a notion closely related with competence.

Subskill difficulty and cognitive levels

The CDM approach provides powerful means to obtain estimations of subskill difficulty (from attribute mastery prevalence) even though more than one subskill may be assessed in one item, which makes comparison between subskill difficulty and cognitive levels more specific and straightforward. The inconsistency between subskill difficulty and cognitive levels was found in this study. Intent and Summary, highest in attribute mastery prevalence (easiest), were located at the higher end of the cognitive level continuum. Moderate in attribute mastery prevalence, both Infer and Style were located at the higher end of the cognitive level continuum. Meaning, lowest in attribute mastery prevalence (most difficult), were located in the middle of the cognitive level continuum. Such inconsistency was often observed with regard to the subskills concerning the understanding of the gist (e.g., Alderson & Lukmani, 1989; Jang, 2009). The difficulty of Summary will be different when it is defined as generalizing implicitly stated main ideas (Hughes, 2003) instead of identifying main ideas based on explicit information. Summary in this study falls into the latter category and thus is easier.

Conclusion

To explore reading subskill associations when multiple subskills are represented by a single item and exploit potentials in CDM application, this study investigated the associations among reading subskills of the TEM8 sample by adopting the approach of cognitive diagnosis, which is sensitive to the estimation of finer-grained latent variables. There are two major research findings. For one thing, based on the output of the CDM analysis, we can chart a network of reading subskills displaying significant links, link strengths, and link directions. To some extent this may reflect the pathways of mastering reading subskills. For the other, the network of reading subskills established through the CDM approach demonstrates subskill association patterns based on cognitive levels and local/global comprehension, which resonates with major reading process models and reflects the hierarchical and compensatory characteristics of reading subskills. Moreover, the CDM approach also confirmed the inconsistency between subskill difficulty and cognitive levels.

As this study was only based on the TEM8 data of part of the advanced English learners in China, the results may be affected by task effects and some findings may not be applicable to all L2 learners. Methodologically, there is a limitation of not adopting think-aloud protocol in the attribute identifying and Q-matrix validating process. The think-aloud protocol can help researchers better understand the thinking processes of test takers. Moreover, this study only focused on the associations among subskills of reading comprehension. Studies that disclose the relations between those subskills and variables of test-takers' underlying language competence, such as listening, writing, grammar or vocabulary, will be of more academic interest.

Based on the empirically established reading subskill associations, future studies can be carried out to integrate the optimal pathways of mastering reading subskills into teaching plans, to investigate the connections between mastering reading subskills and using reading subskills, to disclose relations between internal subskills and external predictors, and to develop test validation tools catering to CDM statistics. Important future strands of L2 research involving cognitive diagnosis may include the following: integrating cognitive diagnosis with L2 ability research, exploring the thinking process of L2 learning, and promoting the effectiveness of L2 teaching and testing.

Acknowledgments

The data of this research were provided by the Committee of Test for English Majors in China

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This was funded by National Social Science Foundation in China (17BYY101)

ORCID

Huilin Chen http://orcid.org/0000-0001-8040-3472
Yuyang Cai http://orcid.org/0000-0002-0320-4602
Jimmy de la Torre http://orcid.org/0000-0002-0893-3863

References

Agresti, A. (2002). *Categorical data analysis*. Wiley.
Akbay, L., & de la Torre, J. (2020). Estimation approaches in cognitive diagnosis modeling when attributes are hierarchically structured. *Psicothema*, *32*(1), 122–129. https://doi.org/10.7334/psicothema2019.182
Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.
Alderson, J. C., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a Foreign Language*, *5*(2), 253–270.
Anderson, L. W., & Krathwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
Bernhardt, E. B. (2005). Process and procrastination in second language reading. *Annual Review of Applied Linguistics*, *25*, 133–150. https://doi.org/10.1017/S0267190505000073
Bernhardt, E. B. (2011). *Understanding advanced second-language reading*. Routledge.
Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives, handbook I: Cognitive domain*. Longmans, Green.
Cai, Y. (2020). *Examining the interaction among components of English for specific purposes ability in reading: The triple-decker model*. Peter Lang.
Cai, Y., & Kunnan, A. J. (2019). Detecting the language thresholds of the effect of background knowledge on a language for specific purposes reading performance: A case of the island ridge curve. *Journal of English for Academic Purposes*, *42*, 1–13. https://doi.org/10.1016/j.jeap.2019.100795
Chen, J. (2017). A residual-based approach to validate Q-matrix specifications. *Applied Psychological Measurement*, *41*(4), 277–293. https://doi.org/10.1177/0146621616686021

Chen, H., & Chen, J. (2013). Validating G-DINA model in language test diagnosis. *Journal of Psychological Science [Chinese]*, *36*, 1470–1475.

Chen, H., & Chen, J. (2016a). Exploring reading comprehension skill relationships through the G-DINA model. *Educational Psychology*, *36*(6), 1049–1064. https://doi.org/10.1080/01443410.2015.1076764

Chen, H., & Chen, J. (2016b). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly*, *13*(3), 218–230. https://doi.org/10.1080/15434303.2016.1210610

Chen, J., 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

Davey, B. (1988). Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers. *The Journal of Experimental Education*, *56*(2), 67–76. https://doi.org/10.1080/00220973.1988.10806468

Davis, F. B. (1968). Research in comprehension in reading. *Reading Research Quarterly*, *3*(4), 499–545. https://doi.org/10.2307/747153

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

de la Torre, J., van der Ark, L. A., & Rossi, G. (2018). Analysis of clinical data from a cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, *51*(4), 281–296. https://doi.org/10.1080/07481756.2017.1327286

Dickinson, E., & Adelson, J. (2016). Choosing among multiple achievement measures: Applying multitrait–multimethod confirmatory factor Analysis to State Assessment, ACT, and student GPA data. *Journal of Advanced Academics*, *27*(1), 4–22. https://doi.org/10.1177/1932202X15621905

Du, W., & Ma, X. (2021). Probing what's behind the test score: Application of multi-CDM to diagnose EFL learners' reading performance. *Reading and Writing*, *34*(6), 1441–1466. https://doi.org/10.1007/s11145-021-10124-x

English Major Division of National Foreign Languages Advisory Board. (2000). *English teaching syllabus for English majors [Chinese]*. Foreign Language Teaching and Research Press.

Fan, T., & Yan, X. (2020). Diagnosing English reading ability in *Chinese* senior high schools. *Studies in Educational Evaluation*, *67*, 100931. https://doi.org/10.1016/j.stueduc.2020.100931

Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2000). An NCME instructional module on exploring the logic of Tatsuoka's rule-space model for test development and analysis. *Educational Measurement: Issues and Practice*, *19*(3), 34–44. https://doi.org/10.1111/j.1745-3992.2000.tb00036.x

Gilbert, N. (1993). *Analyzing tabular data: Loglinear and logistic models for social researchers*. UCL Press.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Literacy Research and Instruction*, *6*(4), 126–135 https://doi.org/10.1080/19388076709556976.

Gough, P. B. (1972). One second of reading. *Visible Language*, *6*(4), 291–320.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign.

Hartz, S., & Roussos, L. (2008). The fusion model for skills diagnosis: Blending theory with practicality. *ETS Research Report Series*, *2*(2), i–57. https://doi.org/10.1002/j.2333-8504.2008.tb02157.x

Heaton, J. B. (1991). *Writing English language tests*. Longman.

Hong, H., Wang, C., Lim, Y. S., & Douglas, J. (2015). Efficient models for cognitive diagnosis with continuous and mixed-type latent variables. *Applied Psychological Measurement*, *39*(1), 31–43. https://doi.org/10.1177/0146621614524981

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153–178). Ablex Publishing Corporation.
Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.
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), 031–073. https://doi.org/10.1177/0265532208097336
Javidanmehr, Z., & Anani Sarab, M. R. (2019). Retrofitting non-diagnostic reading comprehension assessment: Application of the G-DINA model to a high stakes reading comprehension test. *Language Assessment Quarterly*, *16*(3), 294–311. https://doi.org/10.1080/15434303.2019.1654479
Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. https://doi.org/10.1111/lang.12034
Kim, A. H. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, *32*(2), 227–258. https://doi.org/10.1177/0265532214558457
Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel. *Research Branch Report*, 8–75. Chief of Naval Technical Training: Naval Air Station Memphis.
Kintsch, W. (1988). The role of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163
Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes*, *39*(2), 125–128. https://doi.org/10.1080/0163853X.2005.9651676
Koda, K. (1996). L2 word recognition research: A critical review. *Modern Language Journal*, *80*(4), 450–460. https://doi.org/10.1111/j.1540-4781.1996.tb05465.x
Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.
Lee, Y., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to StyleL reading and listening assessments. *Language Assessment Quarterly*, *6*(3), 239–263. https://doi.org/10.1080/15434300903079562
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. https://doi.org/10.1111/j.1745-3984.2004.tb01163.x
Li, F. (2008). *A modified higher-order DINA model for detecting differential item functioning and differential attribute functioning* [Unpublished doctoral dissertation]. University of Georgia.
Liao, C., Kuo, B., Deenang, E., & Mok, M. (2016). Exploratory and confirmatory factor analyses in reading-related cognitive component among grade four students in Thailand. *Educational Psychology: Cognitive Diagnostic Assessment*, *36*(6), 1102–1114. https://doi.org/10.1080/01443410.2015.1058342
Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, *33*(3), 391–409. https://doi.org/10.1177/0265532215590848
Liu, J., & Wu, S. (2019). *Research on inventory of China standards of English [Chinese]*. Higher Education Press.
Lohman, D. F., & Ippel, M. J. (1993). Cognitive diagnosis: From statistically based assessment toward theory-based assessment. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 41–71). Erlbaum.
Luebke, S., & Lorié, J. (2013). Use of Bloom's taxonomy in developing reading comprehension specifications. *Journal of Applied Testing Technology*, *14*, 1–26.
McNeil, L. (2012). Extending the compensatory model of second language reading. *System*, *40*(1), 64–76. https://doi.org/10.1016/j.system.2012.01.011
Moore, B., & Stanley, T. (2013). *Critical thinking and formative assessments: Increasing the rigor in your classroom*. Routledge.

Muijselaar, M., Swart, N., Steenbeek-Planting, E., Droop, M., Verhoeven, L., & De Jong, P. (2017). Developmental relations between reading comprehension and reading strategies. *Scientific Studies of Reading*, *21*(3), 194–209. https://doi.org/10.1080/10888438.2017.1278763

Nassaji, H. (2002). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning*, *52*(2), 439–481. https://doi.org/10.1111/0023-8333.00189

Nassaji, H. (2007). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning*, *57*, 79–113. https://doi.org/10.1111/j.1467-9922.2007.00413.x

Paltridge, B. (1996). Genre, text type, and, and the language learning classroom. *ELT Journal*, *50*(3), 237–243. https://doi.org/10.1093/elt/50.3.237

Pearson, P. D., & Raphael, T. E. (1990). Reading comprehension as a dimension of thinking. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 209–240). Routledge.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell Publishing.

Pierce, M., Katzir, T., Wolf, M., & Noam, G. (2010). Examining the construct of reading among dysfluent urban children: A factor analysis approach. *Journal of Literacy Research*, *42*(2), 124–158. https://doi.org/10.1080/10862961003796140

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology*, *38*(10), 1–23. https://doi.org/10.1080/01443410.2018.1489524

Revision Team of TEM8 Syllabus. (2004). *Test syllabus of TEM8 [Chinese]*. Shanghai Foreign Language Education Press.

Rumelhart, D. E. (1977). *Introduction to human information processing*. John Wiley and Sons.

Rumelhart, D. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Erlbaum.

Shermis, M. D., & Long, S. K. (2009). Multitrait-multimethod analysis of FCAT reading and writing or is it writing and reading? *Journal of Psychoeducational Assessment*, *27*(4), 296–311. https://doi.org/10.1177/0734282908328639

Smith, F. (1977). Making sense of reading—and of reading instruction. *Harvard Educational Review*, *47*(3), 386–395. https://doi.org/10.17763/haer.47.3.3508w0540gug23q5

song, M. Y. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, *25*(4), 435–464. https://doi.org/10.1177/0265532208094272

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*(1), 32–71. https://doi.org/10.2307/747348

Tian, D. (2008). *A study of content validity of reading comprehension of TEM-8 from 2005-2007* [Unpublished master's thesis]. Yangzhou University.

Urquhart, A. H., & Weir, C. J. (2014). *Reading in a second language: Process, product and practice*. Routledge.

van Steensel, R., Oostdam, R., & van Gelderen, A. (2013). Assessing reading comprehension in adolescent low achievers: Subskills identification and task specificity. *Language Testing*, *30*(1), 3–21. https://doi.org/10.1177/0265532212440950

von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, *2*, i–35. https://doi.org/10.1002/j.2333-8504.2005.tb01993.x

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 287–307. https://doi.org/10.1348/000711007X193957

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. https://doi.org/10.1111/j.1745-3984.2011.00142.x

Weir, C., & Khalifa, H. (2008). A cognitive processing approach towards defining reading comprehension. *Cambridge ESOL: Research Notes*, *31*, 2–10.

Yang, M., & Liu, J. (2019). China's standards of *English* language ability and business *English* testing and assessment [*Chinese*]. *Foreign Languages in China*, *16*(3), 13–20.

Yi, Y.-S. (2016). Probing the relative importance of different attributes in L2 reading and listening comprehension items: An application of cognitive diagnostic models. *Language Testing*, *34*(3), 337–355. https://doi.org/10.1177/0265532216646141

Zou, S. (2012). *Language testing [Chinese]*. Shanghai Foreign Language Education Press.

Zou, S., Zhang, Y., & Zhou, Y. (2002). The relationships among item types, reading strategies, and test scores in reading tests [Chinese]. *Foreign Language and Their Teaching*, *158*(5), 19–22.