How vocabulary knowledge and grammar knowledge influence L2 reading comprehension: a finer-grained perspective

Huilin Chen[1] · Huan Mei[2]

Abstract

Based on theories on vocabulary knowledge, grammar knowledge, and reading comprehension subcomponents, ten attributes/subskills were defined for 50 items from relevant subtests of TEM4 (Band Four of Test for English Majors in China). Cognitive diagnosis was conducted on the TEM4 data of the randomly sampled 2285 examinees (roughly at the B2 level) through the saturated generalized deterministic inputs, noisy "and" gate (G-DINA) model. The person parameters obtained from cognitive diagnosis served as the basis for simple multiple regression and path analyses for detecting relationship patterns. The study discovered that the relationship pattern at both construct and attribute/subskill levels can be better described as a mediation pattern in which vocabulary knowledge and its attributes are more suitable to serve as the starting point for reading comprehension. The study also discussed the patterns of the impact of vocabulary and grammar attributes on reading subskills as well as the internal subskill relationships within the construct of reading comprehension.

Keywords Vocabulary knowledge · Grammar knowledge · Reading comprehension · Cognitive diagnosis · Mediation

Introduction

Reading comprehension (RC) is a complex process involving a wide range of linguistic as well as non-linguistic skills and knowledge such as the abilities to "recognize words rapidly and efficiently, develop and use a very large recognition vocabulary, process sentences

✉ Huilin Chen
chlwilliam@hotmail.com

Huan Mei
meihuan2017@126.com

[1] School of Education, Shanghai International Studies University, 550 Dalian Rd. (W), Shanghai, China

[2] School of English Studies, Shanghai International Studies University, 550 Dalian Rd. (W), Shanghai, China

in order to build comprehension, engage a range of strategic processes and underlying cognitive skills, interpret meaning in relation to background knowledge, interpret and evaluate texts in line with reader goals and purposes, and process texts fluently over an extended period of time" (Grabe, 2014, p.8). To gain a better understanding of the factors affecting second-language (L2) RC, a plethora of empirical studies have been conducted, based on which Jeon and Yamashita (2014) found out ten key reading correlates. After a detailed meta-analysis, they further discovered that vocabulary knowledge (VK) and grammar knowledge (GK) were the two strongest ones in terms of the magnitude of correlations with L2 RC. In order to further find out which of the two constructs can promote RC more effectively, researchers (e.g., Raeisi-Vanani & Baleghizadeh, 2022; Zhang, 2012) have recently conducted studies to compare the direct contribution of VK or GK to RC. However, those studies rarely took into consideration the moderating and mediating roles of VK and GK.

Moreover, limited implications for teaching and learning can be drawn from previous research on the relationships among the three general constructs, which calls for a further exploration into the relationships among the subcomponents of the three constructs so as to discover specific teaching and learning strategies. To this end, it is important to employ new tools or methods sensitive to distinguish fine-grained attributes or skills. Cognitive diagnosis, an emerging assessment approach, has the capability of measuring an individual's knowledge structure and the cognitive processes at the level of fine-grained attributes (Gierl et al., 2000). Although cognitive diagnosis studies have explored the subcomponents of RC (Chen et al., 2023; Jang, 2009; Ravand & Robitzsch, 2018), VK (Fan et al., 2018), and GK (Yi, 2017), none of them attempted to establish links among the fine-grained subcomponents of the three constructs.

Therefore, this study intends to find out how VK and GK influence RC when moderating or mediating effects are considered and based on the general relationships among constructs further explore how the subcomponents of VK and GK influence the subcomponents of RC. Investigating the relationships at a fine-grained level may have more implications for establishing specific and practical strategies of teaching and learning concerning the three constructs.

Literature review

Multidimensionality of the three constructs

VK has long been considered to be closely associated with RC (Alderson, 2000; Bernhardt, 2011; Grabe, 2009). Even so, it is not fully understood how different aspects of VK are related to reading and the strength of these associations remains unclear (Cheng & Matthews, 2018). Reasons are twofold. For one thing, although VK has often been viewed as multi-componential (Nation, 2001; Read, 2000; Schmitt, 2014), the test used in many empirical studies was still a general measure. For another thing, presumably, because there was more than one well-established VK framework (Chapelle, 1998; Nation, 1990, 2001; Qian, 2002), challenges were posed for researchers in choosing the most appropriate one to operationalize VK. Accordingly, an array of VK subcomponents have occurred such as vocabulary size, collocation knowledge, morphological knowledge, and word association knowledge (Ma & Lin, 2015; Kieffer & Lesaux, 2008), to name but a few. Despite the ready availability of numerous VK frameworks, it is suggested, however, that the applicability of these frameworks should be evaluated in the research process.

In a similar vein, the importance of GK to RC has long been recognized (e.g., Alderson, 1993; Bernhardt, 2011; Bowey, 1986) while "the specifics of grammatical processes in L2 reading comprehension have remained unclear" (Zhang, 2012, p.560). According to Purpura (2004), GK generally consisted of form and meaning, both of which could be further divided into phonological, lexical, morphosyntactic, cohesive, information management, and interactional levels. While in some second language acquisition studies (e.g., Ellis, 2005, 2006; Han & Ellis, 1998), GK was mostly categorized into explicit and implicit knowledge. It is true that there are no absolute standards for the categorization of GK, yet as Jung (2009) pointed out that "a clear operationalization of grammar was imperative to isolate the contribution made by grammar from that of vocabulary, if possible" (p. 42).

Compared to the taxonomies of the other two constructs, those of RC seem to be much more diversified (Alderson & Lukmani, 1989; Davis, 1968; Heaton, 1991; Munby, 1978), leading to a great deal of subcomponents and subsequently more sophisticated interrelationships among them. For example, Munby defined a total of 19 RC subskills which were shown to be compensatory (McNeil, 2012; Stanovich, 1980) or non-compensatory (Sadoski & Paivio, 2007). One important reason for that may have to do with the greater complexity of the reading process which could be demonstrated by various reading models such as bottom-up (Gough, 1972; Kintsch, 1988, 2005), top-down (Goodman, 1967; Smith, 1977) and interactive model (Rumelhart, 1977). In spite of a large number of RC subskills, most of them, as Fan and Yan (2020) illustrated, could fall into the category of the revised taxonomy of educational objectives (Anderson et al., 2001), namely, remembering, understanding, applying, analyzing, evaluating, and creating.

Relationships among the three constructs

As the importance of VK and GK to RC has been universally acknowledged, quite a few researchers began to include them within the same study to explore the relationships among the three. Among those studies, correlation analysis was sometimes conducted to explore whether VK or GK was significantly correlated with RC and how strong such correlation was. Also, multiple regression or structural equation modeling was often used as the main method of analysis to assess the relative contribution of VK and GK to RC. As for the research results, some studies showed that VK played a more significant role (e.g., Guo & Roehrig, 2011; Joh, 2004; Nassaji, 2003; van Gelderen et al., 2004; Zhang, 2012; Zhang & Koda, 2013) while some others draw conclusions in an opposite way (e.g., Jeon, 2012; Kim & Cho, 2013; Nergis, 2013; van Gelderen et al., 2003). Considering the inconclusive research results, further studies need to be carried out to investigate the relative contribution of VK and GK to RC.

The relationships between VK and GK were also well attested. Although most studies confirmed a strong relationship between the two constructs (e.g., Devescovi et al., 2005; Mariscal & Gallego, 2012; Thordardottir et al., 2002), there are mainly two competing views about the directionality in the relationships. According to the nativist theories (Chomsky, 1981; Baker, 2005), grammar is independent of vocabulary in language development. According to lexicalist theories, in contrast, grammatical structure emerges from generalizations on the basis of lexical items (Tomasello, 2003) and their representations are directly related to more concrete lexical structure (Bannard et al., 2009).

There is also a need to explore the associations among the subcomponents of the three constructs as stated by Choi and Zhang (2021) that "the different dimensions of a type of linguistic knowledge (e.g., vocabulary knowledge) could be analyzed concurrently, together

with the other type (e. g., grammar knowledge), to provide a more complete understanding about how they each and collectively would predict reading comprehension" (p.23). Although some studies (e.g., Raeisi-Vanani & Baleghizadeh, 2022; Zhang, 2012) tried to divide each construct into more specific subcomponents and then explored the relationship among them, they used one test to identify one subcomponent. The problem is that it is almost impossible that a test assessing a single subcomponent exists and such tests cannot reflect the reality of learning practice. In order to identify multiple subcomponents or subskills represented in one test or in one test item, new methods (e.g., cognitive diagnosis) which are sensitive to distinguish finer-grained attributes or subskills need to be adopted.

Compared with studies on the direct contribution of VK and GK to RC, little attention has been given to the moderating and mediating relationships among the three, let alone among their subskills. Admittedly, there were studies (e.g., Joh 2004; Lee, 2016; Shiotsu & Weir, 2007; Taşçı & Turan, 2021) touching upon the issue of moderating effects. However, what they addressed was whether or how such factors as L2 proficiency levels or task types moderated the relationship between VK/GK and RC rather than how VK/GK influenced RC when GK/VK served as moderators. Similarly, there were few studies (e.g., Deacon & Kieffer, 2018; Droop & Verhoeven, 2003; Silva & Cain, 2015) exploring whether or how VK or GK had an indirect effect on RC mediated by factors other than GK and VK. For example, in Deacon and Kieffer's study, path analysis was used to test whether there were indirect associations of GK (syntactic awareness) on RC via word reading or not. It is true that external factors should be reckoned with in exploring the relationship among the three constructs, yet the moderating and mediating relationships among the three constructs (and their attributes) themselves are a more central issue to tackle, which can help us gain a thorough understanding of the associations among the three constructs.

Cognitive diagnosis

Integrating cognitive psychology with psychometrics, cognitive diagnosis is an assessment approach used to measure an individual's knowledge structure and the processing of fine-grained subskills (Gierl et al., 2000; Sawaki et al., 2009). According to Lee and Sawaki (2009a), two major components are included in cognitive diagnosis, one is the Q-matrix identifying the relationships between test items and cognitive attributes (subskills). The other one is psychometric modeling of these attributes and items achieved by cognitive diagnostic models (CDMs). By inputting the Q-matrix and examinees' response data into certain CDM, cognitive diagnostic information can be generated including the overall mastery profile of the examinee group and each individual's mastery probabilities (latent person parameters) on each attribute, etc.

So far, a number of studies have used cognitive diagnosis to extract fine-grained information about students' RC (e.g., von Davier, 2008; Lee & Sawaki, 2009b; Jang, 2009; Wang & Gierl, 2011; Kim, 2015). However, scant attention has been paid to GK and VK in this regard. Among the few studies, Henson and Templin (2007) and Yi (2017) used the same data but different CDMs to diagnose examinees' mastery on three GK subskills. Fan et al. (2018) focused on dyslexic students and identified their strengths and weaknesses in seven VK attributes. Among the previous cognitive diagnosis studies, Lin et al. (2018) seemed to be the only one that involved all the three constructs within one study; however, the relationships among the subskills were not investigated, which then served as the starting point for the present research.

One of the future trends of CDM application in language studies is to address theoretical issues by investigating the relationships between pertinent language skills and knowledge to gain a more comprehensive understanding of the components of language ability (Mei & Chen, 2022).

The present research

As reviewed so far, it has been clear that the three constructs were all multidimensional. Despite the abundant studies that explored the associations among them, the relative importance of VK and GK in RC was still a controversial issue. Moreover, two research areas remained underdeveloped, one was the moderating or mediating roles of VK and GK as the two constructs jointly influence RC and the other one was the effects of VK attributes and GK attributes on RC subskills. In the investigation into the first research area, path analysis at the construct level was carried out to explore the moderating and mediating relationships among the three constructs. The theory-based subcomponents of the three constructs serve as the basis of validating the classifications of attributes and subskills in the three subtests. In the investigation into the second research area, cognitive diagnosis was employed to generate information about individuals' mastery probabilities on each attribute of the three constructs, based on which path analysis at the attribute/subskill level was conducted to explore how VK attributes and GK attributes influence RC subskills. Separate studies may be conducted when VK and GK shift their roles as the moderator or the mediator. In doing so, this study aimed to demystify the complexity of the associations among the three constructs from a more comprehensive and nuanced perspective. To this end, three research questions were addressed:

1. How do VK and GK influence RC when moderating or mediating effects are considered?
2. How do the attributes of VK influence the subskills of RC when GK serves as the moderator or the mediator?
3. How do the attributes of GK influence the subskills of RC when VK serves as the moderator or the mediator?

Material and methods

Data description

This study utilized the vocabulary subtest, the grammar subtest, and the reading comprehension subtest that were administered in the Band 4 Test for English Majors (TEM4) in China. TEM4 is typically given to second-year college students majoring in English in China to evaluate their mastery of the halfway benchmarks of the program. According to Liu and Wu (2019) and Yang and Liu (2019), TEM4 corresponds to B2 levels of CEFR (Common European Framework of Reference for Languages). TEM4 is a mandatory requirement for receiving a bachelor's degree in English majors in China, and thus, it attracts a large number of participants who are high-quality English learners from various regions in China. Therefore, TEM4 candidates were chosen as the participants for this research.

According to Syllabus for English Majors (English Major Division of National Foreign Languages Advisory Board, 2000), TEM4 requires candidates to recognize meanings of

5500–6000 words and accurately and flexibly use 3000–4000 words and their collocations for the vocabulary subtest, requires candidates to master basic grammatical and syntactic rules for the grammar subtest, and requires candidates to master subskills of grasping the general idea of the reading material, understanding the facts and details that explain the general idea, conducting judgment and reasoning, and understanding the logic of the context for the reading comprehension subtest.

As a nationwide high-stake test, TEM4 has long become the target of validation studies. Studies confirmed the good validity for its vocabulary subtest (Peng & Zou, 2012; Treffers-Daller & Huang, 2020), its grammar subtest (Peng & Zou, 2012), and its reading comprehension subtest (Treffers-Daller & Huang, 2020; Yang, 2011). In this study, the original TEM4 subtests and their items were adopted and they were deemed to be valid. The vocabulary subtest includes 15 multiple-choice questions that are scored as either correct or incorrect. These questions assess an individual's ability to differentiate between words that have similar spellings, identify subtle differences in word meanings, and recognize appropriate word combinations within a given context. The grammar subtest also contains 15 multiple-choice questions. These questions assess a wide range of grammatical structures such as subject-predicate agreement, tense, voice, and mood. The reading comprehension subtest contains four articles and 20 multiple-choice items. The four articles are an article about the change of human behaviors caused by mobile phones, a report of social class changes in Britain, an excerpt of the novel *Jane Eyre*, and an article about the function of computers as human companions, respectively. Each article has a length of approximately 400 words and has an average Flesch Kincaid Grade Level Readability (Kincaid et al., 1975) of 8.8. This corresponds to a reading level of around grades 8 to 9 for native English speakers.

This study's participants were randomly selected from 236,586 sophomore English major students across the country who took the same Test for English Majors (TEM4). To select the participants randomly, a selection proportion of approximately 1% was set using SPSS case selection functionality. A total of 2285 TEM4 candidates were chosen via random sampling method. The TEM Examination Board provided the examinee data for this research.

Attribute/subskill definition

The Syllabus for English Majors (English Major Division of National Foreign Languages Advisory Board, 2000) only gives general descriptions about the skill requirements for TEM4, so the attribute/subskill defining process in this study is mainly a combination of analyzing item content and consulting previous attribute/subskill classifications for the three constructs. This research invited three teachers possessing doctoral degrees in Applied Linguistics and two doctoral students who were studying English Linguistics to decide what attributes/subskills were involved in the test under study based on previous attribute/subskill classifications.

After a content analysis of the vocabulary subtest to judge whether the test content suits the theories, the five experts all agreed that Nation's (2001) tripartite classification—Form, Meaning, and Use—was quite applicable to defining the VK attributes in the vocabulary subtest. The experts classified the VK in the subtest as Lexical Form, Lexical Meaning, and Lexical Collocation. In this classification, Nation's Use was renamed as Lexical Collocation in that the subtest focused on recognizing the correct use of collocations.

As for the GK attributes, the experts agreed that Purpura's (2004) framework of GK was quite suitable for this study. According to Purpura (2004), GK generally comprises two major components—form and meaning. Since morphosyntactic forms were the focus of the TEM4 grammar subtest where forms at other levels were not assessed, the experts further classified morphosyntactic forms into Morphological Form and Syntactic Form. Grammatical Meaning was still defined as an attribute because it was indispensable to GK. Therefore, three GK attributes (Morphological Form, Syntactic Form, and Grammatical Meaning) were defined for the grammar subtest.

As for the RC subskills, it was agreed that the taxonomy of thinking skills (Anderson et al., 2001) could be taken as the basis of defining RC subskills because RC subskills are usually related with thinking skills, namely, remembering, understanding, applying, analyzing, evaluating, and creating, as Fan and Yan (2020) have stated. Based on the thinking skills, four RC subskills, namely, Recognizing, Understanding, Inferring, and Evaluation, were established for the reading subtest. Recognizing corresponds to the thinking skill remembering. Inferring here is most closely related with the thinking skill analyzing. Applying and creating were not adopted here because they were not assessed in the reading subtest. This classification of reading subskills also resonates with Luebke and Lorié's classification (2013) which represents a hierarchy of reading subskills on the basis of cognitive levels.

The classifications of the attributes and subskills are largely based on relevant theories on VK, GK, and RC. Those classifications, therefore, did not involve development of new constructs. The definitions of the ten attributes/subskills covering the three constructs are displayed in Table 1. In Appendix A, one item example is provided for each attribute/subskill though it is not necessarily the only attribute/subskill assessed by that item.

G-DINA model

Since CDMs are sensitive to distinguish fine-grained attributes or skills, the researchers utilized the CDM approach to estimate individuals' mastery probabilities for each attribute or subskill, based on which subsequent path analysis was conducted to examine the relationships among VK attributes, GK attributes, and RC subskills.

Several studies, such as von Davier (2008), Lee and Sawaki (2009a, 2009b), Jang (2009), Wang and Gierl (2011), and Kim (2015), have used various CDMs to diagnose reading comprehension subskills. Some scholars have recently suggested that a saturated CDM should be used due to unclear associations between these subskills, as it can accommodate various interactions among them, as noted by Li et al. (2016). Several studies (e.g., Ravand & Robitzsch, 2018; Chen & Chen, 2013, 2015, 2016) have applied the saturated Generalized Deterministic Inputs, noisy "and" gate (G-DINA) model to diagnosing subskills related to reading. These studies have found that the model is a valid and practical tool for this purpose.

The formula of the saturated G-DINA model (de la Torre, 2011) is as follows:

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

The detailed description of the formula can be found in the research about the development of the model (de la Torre, 2011). Earlier research (Ravand & Robitzsch, 2018; Chen and Chen, 2013, 2015, 2016) has shown that the G-DINA model is able to provide

Table 1 Attribute/subskill definitions

Construct	Attribute/subskill	Codename	Definition
VK	Lexical Form	LF	Identifying words according to rules of word formation
	Lexical Meaning	LM	Knowing the conceptual and extended meanings of words
	Lexical Collocation	LC	Knowing appropriate word collocations according to context
GK	Morphological Form	MF	Mastering grammatical categories related with morphological variations
	Syntactic Form	SF	Mastering grammatical categories related with linking components of simple sentences and complex sentences
	Grammatical Meaning	GM	Mastering the use of grammar according to contextual meaning
RC	Recognizing	RG	Identifying details in a text by recalling the information explicitly stated in the reading items
	Understanding	UD	Understanding facts and ideas by organizing, comparing, interpreting, giving descriptions, and stating main ideas
	Inferring	IF	Making inference of implicit information according to context
	Evaluating	EV	Making judgments about information, validity of ideas, or quality of work based on a set of criteria

a satisfactory fit to reading assessment. Therefore, the current study employs the G-DINA model. The CDM analysis was performed using the G-DINA analysis program based on OX Edit software (de la Torre, 2011).

Path analyses

An important output of CDM analysis is the person parameter estimates indicating each individual's probability of mastering each attribute/subskill. Based on the person parameter estimates, the study conducted a series of multiple regression analyses and path analyses. First, multiple regression, which is the basis of path analysis, was conducted to find out the global impact of the VK construct and GK construct on the RC construct and also estimate the impact of specific attributes/subskills of VK and GK on the RC construct. Since those relationships only involve direct effects and there is no need to display their simple path diagrams, multiple regression instead of path analysis was employed here. Then, this research adopted path analyses, which have capability to chart diagrams for complex relationships, to explore whether there existed moderation and mediation relationships among the three constructs globally. Path analyses were conducted again to see how the relationships operated among the attributes/subskills so that specific implications could be derived if strong moderating and/or mediating effects were discovered. In this study, this research adopted AMOS to analyze moderation and mediation effects through path analysis.

Results

Q-matrix validation

The five experts independently coded the attributes/subskills for the items adopted in this study. The coding for vocabulary, grammar, and reading subtests was based on the attributes/subskills of the three constructs respectively because the coding process only focused on the central attributes/subskills of an item assessing a certain construct. A survey was conducted to check if experts agreed on the coding of each attribute/subskill for each item. If at least three out of five experts agreed on a certain attribute/subskill coding for an item, the coding was considered valid for that item. If fewer than three agreed, it was considered invalid. Using this method, a coding matrix was established with 85.8% average agreement among the experts, meaning they agreed on most of the attribute/subskill codings.

Following the initial coding of subskills, a series of validation estimations were performed through the built-in validation function of the G-DINA analysis program and modifications were made accordingly. To identify the misspecifications at the item level, calculation of two statistics was involved: (1) calculating the residue between the observed and predicted correlation of item pairs using Fisher's transformation (the r statistics) and (2) calculating the residue between the observed and predicted log-odds ratios of pairwise item responses (the l statistics). By averaging the z-scores of residuals for each specific item, the values of sr_j and sl_j were obtained. The item with the highest sr_j or sl_j value was deemed to be the item with the most likely misspecification. Based on consensus among experts, the coding of the identified misspecified item was adjusted in each validation estimate. The CDM analysis's absolute model fit at the test level was then updated after each modification. Chen (2017) provided further details regarding this process. The procedure continued until the model fit was satisfactory.

Using the r or l statistics, the overall fit of the model at the test level was estimated by testing the null hypothesis that the residuals are not statistically significant (i.e., approaching 0). According to Chen, de la Torre, and Zhang (2013), the model is considered to have a satisfactory fit if the maximum z_r or z_l statistic is smaller than the Bonferroni adjusted critical z-score z_c at a certain significance level.

Following several adjustments, the absolute model fit at the test level was eventually deemed significant at a p-value of 0.05, which equates to the Bonferroni adjusted $z_c = 4.10$. The maximum z_r and z_l values were found to be 4.05 and 4.09, respectively. That not only demonstrated that the G-DINA model, in combination with the coding matrix, was able to provide an appropriate fit for the test data but also provided evidence that the test had good construct validity in terms of the composition of attributes. Though the validation process continued, the researchers were unable to make any further improvements regarding the model fit.

The coding matrix that was obtained was shared with the coders again to seek their opinions. After further discussion, they reached a consensus on the final Q-matrix, which can be found in Appendix B.

Person parameter estimates

The person parameter demonstrates the posterior probability of each individual's mastery of each attribute/subskill. It was used to represent the level of mastery that an individual had achieved for the specified attribute/subskill. Appendix C includes the estimates for the first 23 participants from the working data, which represents 1% of the total sample. Person parameter estimates serve as the basis of the following analyses.

Simple regression analyses

As an initial investigation into Research Question 1, multiple regression was first conducted to find out how VK and GK directly influenced RC at the construct level. The multiple regression was realized by using SPSS. The scores of the three constructs for each participant were obtained by averaging that participant's person parameter estimates concerning relevant constructs so as to ensure that all calculations were based on person parameter estimates. Table 2 shows the effects in the simple model at the construct level established through simple multiple regression.

Table 2 shows that the combination of VK and GK significantly influenced RC ($F = 1740.817$, $p < 0.01$), with both VK and GK significantly influencing RC ($p < 0.01$). The beta weights suggested that the effect of GK on RC was slightly larger than that of VK on

Table 2 Effects in the construct level simple model

Variables		R	R^2	Adjusted R^2	F	Beta	t
Dependent variable	RC	0.78	0.60	0.60	1740.82**		
Independent variables	VK					0.40	14.32**
	GK					0.41	14.73**

**p < 0.01

RC. The R square value was 0.604, which indicated that 60.4% of the variance in RC was explained by the model.

In order to discover the impact of VK and GK on RC at an attribute level instead of at the construct level, another multiple regression was conducted. The six variables for VK/GK attributes were directly provided in person parameter estimates and the scores of RC for each participant were obtained by averaging that participant's person parameter estimates concerning the RC construct. Table 3 shows the effects in the simple model at the attribute level established through simple multiple regression.

Table 3 shows that the combination of six VK/GK attributes significantly influenced RC ($F = 601.292$, $p < 0.01$), with all VK/GK attributes except LM significantly influencing RC ($p < 0.01$). The beta weights suggested that the effects of SF and MF on RC were much larger than those of other attributes on RC. The R square value was 0.612, which indicated that 61.2% of the variance in RC was explained by the model.

Moderation and mediation analyses at the construct level

However, the above regression analyses only demonstrate the direct effects of VK and GK on RC. The focus of Research Question 1 lies in whether VK or GK influenced RC indirectly. Therefore, there is a need to conduct moderation and mediation analyses to find out such indirect effects. Moderation and mediation analyses at the construct level were first conducted to see whether there were moderation or mediation relationships globally.

Path analyses were conducted with AMOS to explore whether there existed moderation and mediation relationships among the three constructs globally. The scores of the three constructs for each participant were obtained by averaging that participant's person parameter estimates concerning relevant constructs.

Moderation relationship in essence is represented by the linear interaction between two independent variables and the coefficient of the interaction variable. Figure 1 shows the moderation path diagram in the saturated model.

In order to find out how large R^2 changed (not shown in AMOS) after adopting the moderation model, corresponding hierarchical multiple regression analysis was conducted by means of SPSS and the same results together with the R^2 change were obtained. All the results are displayed in Table 4.

In Table 4, the coefficient of VK × GK was significant ($p < 0.01$), which demonstrates that there was a moderation effect when VK or GK was regarded as a moderator. However, given the very small value of R^2 change, the moderation model could not be

Table 3 Effects in the attribute level simple model

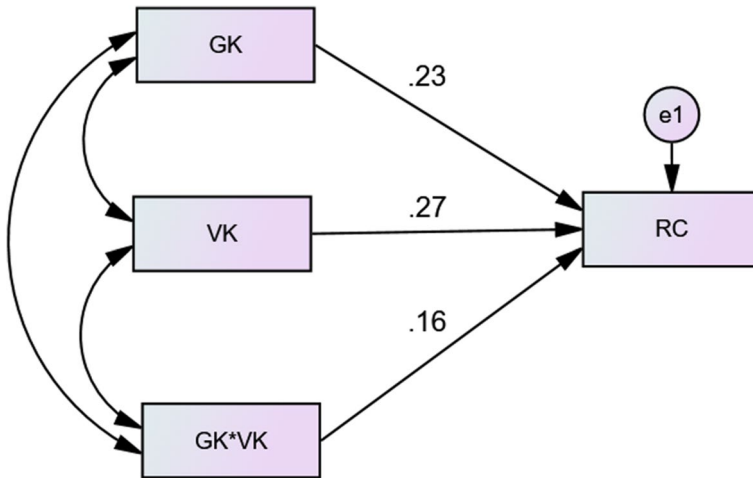| Variables		R	R^2	Adjusted R^2	F	Beta	t
Dependent variable	RC	0.78	0.61	0.61	601.29**		
Independent variables	LF					0.19	9.87**
	LM					0.05	1.87
	LC					0.18	9.29**
	MF					0.21	7.66**
	SF					0.22	9.81**
	GM					0.12	5.12**

**$p < 0.01$

Fig. 1 Construct level moderation model. GK*VK represents the interaction between GK and VK. The effect of GK*VK on RC is regarded as the moderation effect

Table 4 Effects in the construct level moderation model

Variables		R	R^2	F	Beta	t	R^2 change
Dependent variable	RC	0.78	0.61	1171.59**			0.002
Independent variables	VK				0.27	8.07**	
	GK				0.23	6.50**	
	VK × GK				0.16	3.71**	

**$p < 0.01$

a good one because R^2 only increased by 0.002 when the simple model was changed to the moderation model.

The research then conducted path analyses to find out whether the mediation model could be better. Figure 2 shows the mediation path diagrams in saturated models when VK or GK is the mediator.

The indirect effect was calculated by multiplying together the effect of the independent variable on the mediator and that of the mediator on the dependent variable. Table 5 shows the results of mediation analyses at the construct level.

The 95% bootstrap confidence interval was the criteria to determine whether the indirect effect in mediation is significant or not (Hayes, 2018). If the 95% bootstrap confidence interval does not include 0 and the indirect effect falls within it, the indirect effect is significant. Whether VK or GK was taken as the mediator, all the total effects and the direct effects were statistically significant ($p < 0.01$), and all the indirect effects met the 95% bootstrap confidence interval criteria and therefore were statistically significant. In each model, the direct effect was larger than the indirect effect. Compared with the model in which VK was the mediator, all the three effects were larger when GK served as the mediator.

Fig. 2 Construct level mediation models. The upper diagram shows the mediation model with VK as the mediator. The lower diagram shows the mediation model with GK as the mediator
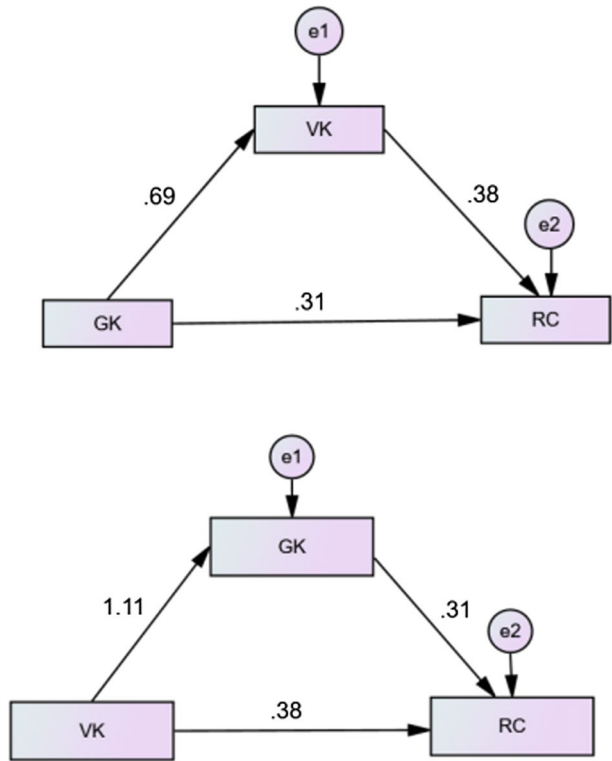
Table 5 Effects in construct level mediation models

Independent variable	Mediator	Total effect	Direct effect	Indirect effect
GK	VK	0.57**	0.31**	0.26 95% BCI, 0.22–0.30
VK	GK	0.72**	0.38**	0.34 95% BCI, 0.29–0.40

**p < 0.01*

BCI bootstrap confidence interval

Mediation analyses at the attribute/subskill level

The above analyses showed that both moderating and mediating effects were discovered at the construct level, which is possible when the phenomenon under investigation is conceptualized and tested from different perspectives (Hayes, 2018). However, mediating effects were found to be much stronger than moderating effects at the construct level. In order to discover a better model, this study only focused on the mediation analysis at the attribute/subskill level so that the implications of this study could be not only specific but also more powerful.

All mediation analyses at the attribute/subskill level were conducted with AMOS to find out how VK/GK attributes influenced RC subskills through VK or GK mediator so as to deal with Research Questions 2 and 3. In other words, the independent variables were VK/GK attributes, the dependent variables were RC subskills, and the mediator was the VK/GK construct variable which was composed of each participant's average score of the person parameter estimates concerning that construct.

When VK served as the mediator, the three GK attributes and the four RC subskills formed a combination of 12 mediating relationships which are displayed in Fig. 3.

By adding all covariances among the VK attributes and some links among the RC subskills, the model fit of the path diagram was achieved (χ^2/df = 0.02, p = 0.88, RMSEA = 0.00). The effects of VK attributes on RC subskills are shown in Table 6.

When GK served as the mediator, the three VK attributes and the four RC subskills formed a combination of 12 mediating relationships which are displayed in Fig. 4.

By adding all covariances among the GK attributes and some links among the RC subskills, the model fit of the path diagram was achieved (χ^2/df = 0.02, p = 0.89, RMSEA = 0.00). The effects of GK attributes on RC subskills are shown in Table 7.

In the process of validating the above mediation analyses, some links were added among RC subskills. Those links demonstrated the relationships among RC subskills and the effects in the two mediation models are compared in Table 8.

In Table 8, the RC subskill relationships were very similar in the two mediation models, which probably demonstrated that there existed a stable internal subskill network in the reading construct.
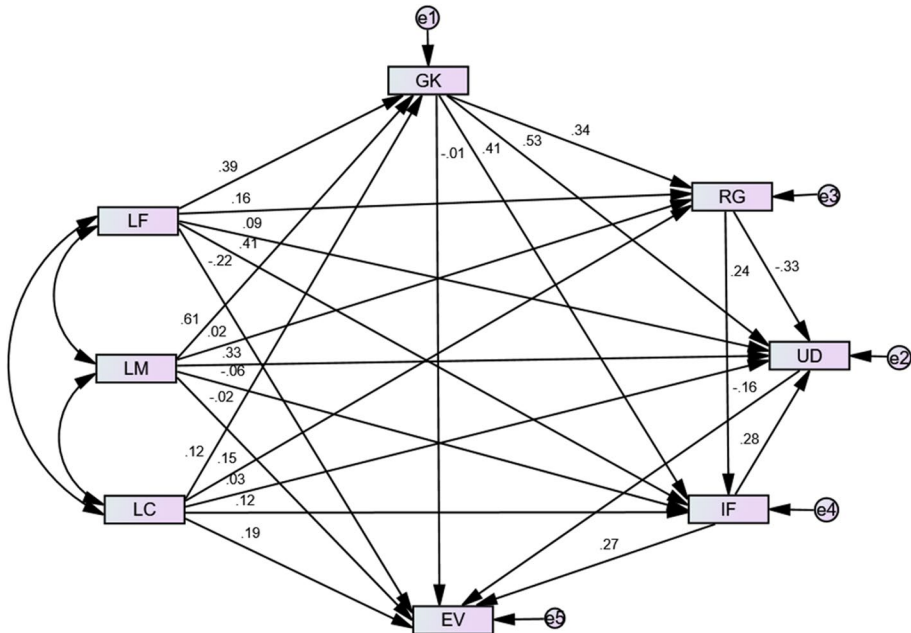

Fig. 3 Attribute/subskill level mediation model (GK as mediator). The figure includes a combination of 12 mediation models with GK as the common mediator. The independent variables are LF, LM, and LC, respectively. The dependent variables are RG, UD, IF, and EV, respectively. The figure also displays the associations among RC subskills

Table 6 Effects in the attribute/subskill level mediation model (GK as mediator)

Independent variable	Mediator	Dependent variable	Total effect	Direct effect	Indirect effect
LF	GK	RG	0.29**	0.16**	0.13**
LM			0.23**	0.02	0.21**
LC			0.19**	0.15**	0.04**
LF		UD	0.39**	0.09**	0.29**
LM			0.65**	0.33**	0.32**
LC			0.09**	0.03	0.06**
LF		IF	0.64**	0.41**	0.23**
LM			0.25**	−0.06	0.31**
LC			0.21**	0.12**	0.09**
LF		EV	−0.11**	−0.22**	0.11**
LM			−0.06	−0.02	−0.05
LC			0.23**	0.19**	0.04**

**$p < 0.01$

Note. The statistical significance of all the indirect effects was also based on the 95% bootstrap confidence interval criteria
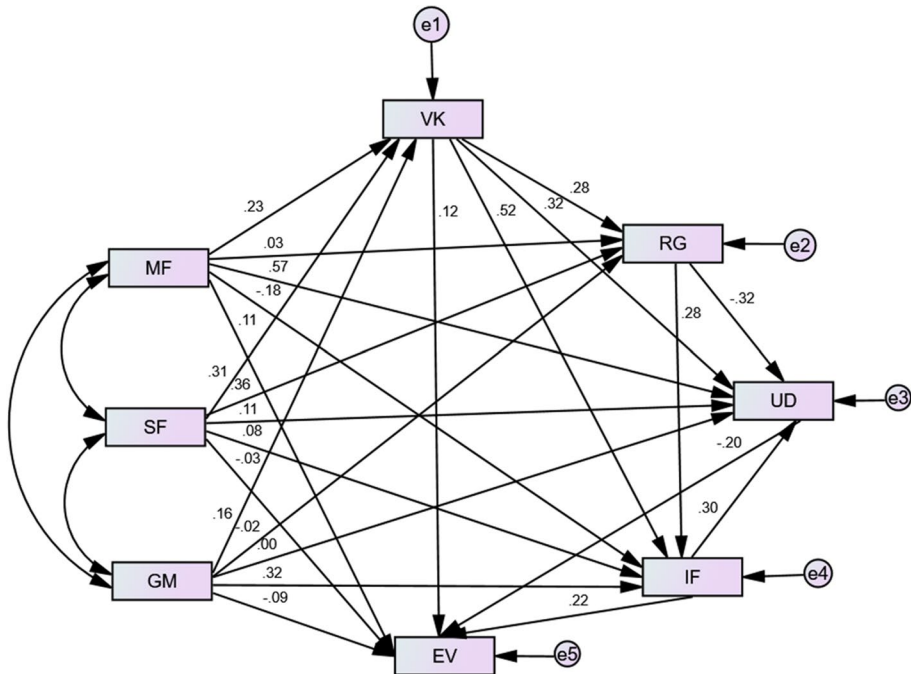

Fig. 4 Attribute/subskill level mediation model (VK as mediator). The figure includes a combination of 12 mediation models with VK as the common mediator. The independent variables are MF, SF, and GM, respectively. The dependent variables are RG, UD, IF, and EV, respectively. The figure also displays the associations among RC subskills

Table 7 Effects in the attribute/subskill level mediation model (VK as mediator)

Independent variable	Mediator	Dependent variable	Total effect	Direct effect	Indirect effect
MF	VK	RG	0.09*	0.03	0.06**
SF			0.45*	0.36*	0.09**
GM			0.02	−0.02	0.04**
MF		UD	0.60*	0.57*	0.04*
SF			0.17**	0.11**	0.07*
GM			0.16**	−0.00	0.17**
MF		IF	−0.03	−0.18*	0.15*
SF			0.37*	0.08	0.29*
GM			0.41**	0.32**	0.09*
MF		EV	0.01	0.11*	−0.10**
SF			0.06	−0.03	0.08*
GM			−0.01	−0.09	0.08**

*$p < 0.05$; **$p < 0.01$

Note. The statistical significance of all the indirect effects was also based on the 95% bootstrap confidence interval criteria

Table 8 Comparison of RC subskill networks

Relationship between RC subskills	Effect in VK-GK-RG model	Effect in GK-VK-RG model
RG → UD	−0.33**	−0.32**
RG → IF	0.24**	0.28**
IF → UD	0.28**	0.30**
UD → EV	−0.16**	−0.20**
IF → EV	0.27**	0.22**

**$p < 0.01$

Discussion

By adopting the G-DINA model, a saturated CDM accommodating all possible attribute/subskill interactions, this study defined multiple attributes/subskills for most items and retrieved the latent person parameters of attribute/subskill mastery. Based on the person parameters, simple multiple regression was conducted to find out how each VK/GK attribute influenced reading comprehension. The attribute level multiple regression provided more detailed information than the construct level regression did. The construct level regression only showed that the effect of GK on RC was slightly larger than that of VK on RC. The attribute level regression, however, demonstrated that SF, MF, LF, LC, and GM ranked from large to small according to their effects on RC and LM failed to have a significant effect on RC. Unlike the roughly equal construct impact of GK and VK, the overall impact of GK attributes more surpassed that of VK attributes. Both construct level and attribute level effects of GK and VK on RC construct coincide with Jeon and Yamashita's (2014) meta-analysis result and generally accord with the research findings that GK plays

a more contributive role in RC (Jeon, 2012; Kim & Cho, 2013; Nergis, 2013; van Gelderen et al., 2003).

Compared with studies on the direct impact of VK and GK on RC, little attention has been given to the roles of GK when VK was influencing RC or the roles of VK when GK was influencing RC. Corresponding to Research Question 1, this study conducted both moderation analysis and mediation analysis and discovered that almost all effects in mediating relationships were much larger than the main effects in moderating relationships. Therefore, the relationships among the three constructs could be better interpreted as a mediating pattern. In the mediating relationships, the direct impact of GK on RC was smaller than that of VK on RC. The results were inconsistent with those of simple multiple regression analyses probably because indirect effects were not taken into account in simple multiple regression analyses. Although both VK and GK could indirectly influence RC, the indirect impact of VK via GK was larger than that of GK via VK, which reflects the lexicalist view that GK is dependent on VK (Bannard et al., 2009; Tomasello, 2003). The cognitive process concerning the two constructs seems to be consistent with the language development route. Therefore, VK is more suitable to serve as the starting point to acquire RC while GK, on the basis of VK, can contribute more to RC in indirect effects.

In dealing with Research Questions 2 and 3, the mediation relationships at the attributes/subskill level were investigated; the magnitudes of direct effects seemed to be larger than those of indirect effects, whereas the indirect effects were found to be dominant when only statistical significance was considered. At the attribute/subskill level, the number of significant indirect effects far surpassed that of significant direct effects. Therefore, the mediation analyses conducted at the attribute/subskill level discovered that the impact of VK and GK on RC may not be fully disclosed through simple regression analyses or even mediation analyses at the construct level because a large number of VK and GK effects are indirect and the larger direct effects at the construct level may cover up the fact that indirect effects are to some extent dominant at the attribute/subskill level. The mediation analyses conducted at the attribute/subskill level again reflect the lexicalist view because, compared with the GK-VK-RC model, the VK-GK-RC model possessed more effects which were larger and significant.

Insights on strategies of improving reading can also be obtained according to the analyses at the attribute/subskill level. When GK served as a mediator, the comparatively weak direct effects from LM (only significant on UD) demonstrate that form instead of meaning is more important to the sample students in the reading process. When VK served as a mediator, the comparatively weak direct effects from GM (only significant on IF) confirm the importance of form to the sample students. The finding may imply that it is more advisable for the sample students to use a somewhat bottom-up strategy in which language forms at lower cognitive levels are the focus in reading. Another finding is that VK and GK attributes tend to have a greater impact on the RC subskills derived from intermediate or higher-order thinking skills instead of those based on lower-order thinking skills probably because the knowledge competence concerning vocabulary and grammar may promote higher-order thinking (Thompson, 2001; Ünaldı & Yüce, 2021) and therefore the learning of the two constructs should be encouraged. Moreover, most VK/GK attributes failed to have significant positive effects on the top-order subskill EV. The only exception was LC. One explanation might be that a large number of lexical collocations are idioms which are closely related with text style. Therefore, the learning of LC might be essential to enhance reading competence at higher cognitive levels.

In the process of exploring the effects of VK and GK on RC, the study also disclosed the internal relationships among reading subskills. In both mediation models at the attribute/

subskill level, patterns of relationships among reading subskills were the same and the effects were all significant and very similar across the two models, which demonstrated that there existed a stable internal subskill network in the reading construct. Insights on strategies of improving reading can also be obtained from the internal subskill relationships. It is advisable to take RG as the starting point of learning to improve UD and IF which in turn can improve EV. The directions of the links among reading subskills roughly represented a bottom-up pattern (Gough, 1972; Kintsch, 1988, 2005; Rumelhart, 1977) in which subskills at lower cognitive levels are the antecedents of those at higher cognitive levels. The study shows that a bottom-up pattern may dominate the reading cognition of TEM4 test takers who are roughly at the B2 level of CEFR probably because less competent readers tend to adopt bottom-up strategies and more competent readers may use top-down or interactive ones (e.g., Cziko, 1980; Geladari et al., 2010). The significant negative effects between two subskills (RG and UD, UD and EV) can be explained with previous research (Bernhardt, 2005; McNeil, 2012; Stanovich, 1980) on compensatory relationships between reading subskills at different cognitive levels. That is to say, when one's competence at UD is low, his/her competence at RG and EV may increase to compensate for the lack of UD, and vice versa.

The detection of a stable reading subskill network in the mediation models discussed above may provide additional evidence that the mediation models, especially the one with GK as a mediator, are good models to accommodate both external and internal relationships.

Conclusion

By adopting cognitive diagnosis, simple multiple regression analyses, and path analyses, this study explored the impact of VK and GK on RC from a finer-grained perspective. Based on theories on the components of the three constructs (Anderson et al., 2001; Fan & Yan, 2020; Luebke & Lorié, 2013; Nation, 2001; Purpura, 2004), ten attributes/subskills were defined for 50 TEM4 items. Cognitive diagnosis was carried out on the TEM4 data of 2285 examinees through the saturated G-DINA model. The person parameters obtained from CDM analysis served as the basis for simple multiple regression and path analyses conducted at both construct and attribute/subskill levels. Simple models were established through simple multiple regression while moderation and mediation models were established through path analyses. Since the moderating effect of the VK or GK construct was very weak, an in-depth study at the attribute/subskill level was not carried out for moderation models.

This study may shed new light on the controversy on whether GK or VK contributes more to RC. The study discovered that the relationship pattern among the three constructs can be better described as a pattern of mediation and VK is more suitable to serve as the starting point to acquire RC while GK, on the basis of VK, can contribute more to RC in indirect effects. It was also discovered that the impact of VK and GK on RC may not be fully disclosed through simple regression analyses or even mediation analyses at the construct level because a large proportion of VK and GK effects are indirect and the larger direct effects at the construct level may cover up the fact that indirect effects are to some extent dominant at attribute/subskill level.

The mediation analyses conducted at the attribute/subskill level may provide insights on strategies of improving reading. For the TEM4 test takers who are roughly at the B2

level of CEFR, VK/GK attributes closely related with form usually have stronger total and direct effects on reading while VK/GK attributes closely related with meaning usually have stronger indirect effects compared with their corresponding direct effects. Similarly, the inadvertently detected internal subskill network of RC was found to have a bottom-up pattern. In other words, superficial forms may prevail over contextual meanings not only in the reading process itself but also in a broader picture of reading for independent English users at the B2 level.

Although this study aims to investigate the impact of VK and GK on RC from a finer-grained perspective, attributes and subskills were not analyzed for moderators and mediators, which may not provide a complete finer-grained picture of the relationships among the three constructs. Although this study explored the influence of the two most important contributors on reading from a finer-grained perspective, it will be more interesting if more reading correlates proposed by Yamashita (2014) can be analyzed from such perspective. Further research is expected to cope with those limitations.

Declarations

Competing interests Huilin Chen and Huan Mei do not have competing financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

References

Alderson, J. C. (1993). The relationship between grammar and reading in an English for academic purposes test battery. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 203–219). Alexandria, VA: Teachers of English to Speakers of Other Languages.

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*, 253–270.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A taxonomy for learning teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Baker, M. C. (2005). Mapping the terrain of language learning. *Language Learning and Development, 1*(1), 93–129. https://doi.org/10.1207/s15473341lld0101_6

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences, 106*(41), 17284. https://doi.org/10.1073/pnas.0905638106

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.

Bowey, J. A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology, 41*, 282–299. https://doi.org/10.1016/0022-0965(86)90041-X

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge University Press.

Chen, J. (2017). A residual-based approach to validate Q-matrix specifications. *Applied Psychological Measurement, 41*, 277–293. https://doi.org/10.1177/0146621616686021

Chen, H., & Chen, J. (2015). 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. (2016). 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*, 123–140. https://doi.org/10.1111/jedm.2013.50.issue-2

Chen, H., Cai, Y., & de la Torre, J. (2023). Investigating Second Language (L2) Reading subskill associations: A cognitive diagnosis approach. *Language Assessment Quarterly, 20*(2), 166–189. https://doi.org/10.1080/15434303.2022.2140050

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

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 3–25. https://doi.org/10.1177/0265532216676851

Choi, Y., & Zhang, D. (2021). The relative role of vocabulary and grammatical knowledge in L2 reading comprehension: A systematic review of literature. *International Review of Applied Linguistics in Language Teaching, 59*(1), 1–30. https://doi.org/10.1515/iral-2017-0033

Chomsky, N. (1981). Principles and parameters in syntactic theory. *Explanation in linguistics: The logical problem of language acquisition,* pp. 32–75.

Cziko, G. A. (1980). Language competence and reading strategies: A comparison of first- and second-language oral reading errors. *Language Learning, 30*(1), 101–114.

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

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

Deacon, S. H., & Kieffer, M. (2018). Understanding how syntactic awareness contributes to reading comprehension: Evidence from mediation and longitudinal models. *Journal of Educational Psychology, 110*(1), 72–86. https://doi.org/10.1037/edu0000198

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language, 32*(4), 759–786. https://doi.org/10.1017/S0305000905007105

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second language learners. *Reading Research Quarterly, 38*(1), 78–103. https://doi.org/10.1598/RRQ.38.1.4

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172. https://doi.org/10.1017/s0272263105050096

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27*, 431–463. https://doi.org/10.1093/applin/aml022

English Major Division of National Foreign Languages Advisory Board. (2000). *English Teaching Syllabus for English Majors [Chinese].* Beijing: 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

Fan, X., Wang, S., Geng, B., Wu, H., & Wang, M. (2018). Cognitive diagnostic tests of Chinese vocabulary for dyslexic students among grade 4 and 5. *Educational Measurement and Evaluation, 3*, 14–19. https://doi.org/10.16518/j.cnki.emae.2018.03.003

Geladari, A., Griva, E., & Mastrothanasis, K. (2010). A record of bilingual elementary students' reading strategies in Greek as a second language. *Procedia-Social and Behavioral Sciences, 2*(2), 3764–3769. https://doi.org/10.1016/j.sbspro.2010.03.585

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

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Literacy Research and Instruction, 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.

Grabe, W. (2014). Key issues in L2 reading development. In X. Deng & R. Seow (Eds.), *Alternative pedagogies in the English language & communication classroom* (pp. 8–18). Centre for English Language Communication, National University of Singapore.

Guo, Y., & Roehrig, A. D. (2011). Roles of general versus second language (L2) knowledge in L2 reading comprehension. *Reading in a Foreign Language, 23*(1), 42–64.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 2*(1), 1–23. https://doi.org/10.1177/136216889800200102

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs, 85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

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

Henson, R., & Templin, J. (2007). *Large-scale language assessment using cognitive diagnosis models*. Paper presented at the annual meeting of the National Council for Measurement in Education, Chicago, IL.

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. https://doi.org/10.1177/0265532208097336

Jeon, E. H. (2012). Oral reading fluency in second language reading. *Reading in a Foreign Language, 24*(2), 186–208. https://doi.org/10.1177/026455059304000405

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

Joh, J. (2004). Which prior knowledge? Relative contribution of prior knowledge components to EFL reading comprehension. *Language Research, 40*(4), 1035–1057. https://doi.org/10.1111/j.1467-8365.1984.tb00141.x

Jung, J. (2009). Second language reading and the role of grammar. *Working Papers in TESOL & Applied Linguistics, 9*(2), 29–48. https://doi.org/10.7916/salt.v9i2.1447

Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing, 21*(8), 783–804. https://doi.org/10.1007/s11145-007-9092-8

Kim, A. Y. (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

Kim, J., & Cho, Y. (2013). Relative roles of vocabulary and grammar knowledge in second language reading. *Korean Journal of English Language and Linguistics, 13*(3), 497–520.

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*, pp. 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*, 163–182.

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

Lee, J. W. (2016). The role of vocabulary and grammar in different L2 reading comprehension measures. *English Teaching, 71*, 79–97. https://doi.org/10.15858/engtea.71.3.201609.79

Lee, Y. W., & Sawaki, Y. (2009). Cognitive diagnosis and Q-matrices in language assessment. *Language Assessment Quarterly, 6*(3), 169–171. https://doi.org/10.1080/15434300903059598

Lee, Y. W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*(3), 239–263. https://doi.org/10.1080/15434300903079562

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

Lin, Y., Chen, H., & Chen, J. (2018). Exploring cognitive diagnosis retrofitting and further analyses of language proficiency testing: The case of the Guangzhou English achievement examination [Chinese]. *Journal of Psychological Science, 41*(4), 989–995.

Liu, J., & Wu, S. (2019). *Research on Inventory of China Standards of English [Chinese]*. Higher Education Press.

Luebke, S., & Lorié, J. (2013). Use of Bloom's taxonomy in developing reading comprehension specifications. *Journal of Applied Testing Technology, 14*, 1–26.

Ma, Y. H., & Lin, W. Y. (2015). A study on the relationship between English reading comprehension and English vocabulary knowledge. *Education Research International*, Article 209154. https://doi.org/10.1155/2015/209154

Mariscal, S., & Gallego, C. (2012). The relationship between early lexical and grammatical development in Spanish: Evidence in children with different linguistic levels. *The Spanish Journal of Psychology, 15*(1), 112–123. https://doi.org/10.5209/rev_SJOP.2012.v15.n1.37293

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

Mei, H., & Chen, H. (2022). Cognitive diagnosis in language assessment: A thematic review. *RELC Journal*, 00336882221122357. https://doi.org/10.1177/00336882221122357

Munby, J. L. (1978). *Communicative syllabus design*. Cambridge university Press.

Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal, 87*(2), 261–276. https://doi.org/10.1111/1540-4781.00189

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, Mass: Newbury House.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nergis, A. (2013). Exploring the factors that affect reading comprehension of EAP learners. *Journal of English for Academic Purposes, 12*(1), 1–9. https://doi.org/10.1016/j.jeap.2012.09.001

Peng, K., & Zou, S. (2012). Investigating the construct validity of the TEM4 grammar and vocabulary test through the Rasch analysis and confirmatory analysis [Chinese]. *Foreign Languages and Their Teaching, 06*, 49–55.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge University Press.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513–536. https://doi.org/10.1111/1467-9922.00193

Raeisi-Vanani, A., & Baleghizadeh, S. (2022). The contributory role of grammar vs. vocabulary in L2 reading: An SEM approach. *Foreign Language Annals, 55*(2), 559–585. https://doi.org/10.1111/flan.12606

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

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

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

Sadoski, M., & Paivio, A. (2007). Toward a unified theory of reading. *Scientific Studies of Reading, 11*(4), 337–356. https://doi.org/10.1080/10888430701530714

Sawaki, Y., Kim, H. J., & Gentile, C. (2009). Q-Matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Language Assessment Quarterly, 6*(3), 190–209. https://doi.org/10.1080/15434300902801917

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913–951. https://doi.org/10.1111/lang.12077

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing, 24*(1), 99–128. https://doi.org/10.1177/0265532207071513

Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology, 107*(2), 321–331. https://doi.org/10.1037/a0037769

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

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

Taşçı, S., & Turan, Ü. D. (2021). The contribution of lexical breadth, lexical depth, and syntactic knowledge to L2 reading comprehension across different L2 reading proficiency groups. *English Teaching & Learning, 45*, 145–165. https://doi.org/10.1007/s42321-020-00065-z

Thompson, M. C. (2001). Vocabulary and grammar: Critical content for critical thinking. *Journal of Secondary Gifted Education, 13*(2), 60–66. https://doi.org/10.4219/jsge-2002-367

Thordardottir, E. T., Weismer, S. E., & Evans, J. L. (2002). Continuity in lexical and morphological development in Icelandic and English-speaking 2-year-olds. *First Language, 22*(1), 3–28. https://doi.org/10.1177/014272370202206401

Tomasello, M. (2003). *Constructing a language: A usage based theory of language acquisition*. Harvard University Press.

Treffers-Daller, J., & Huang, J. (2020). Measuring reading and vocabulary with the Test for English Majors Band 4: A concurrent validity study. In J. Clenton & P. Booth (Eds.), *Vocabulary and the four skills: Pedagogy, practice, and implications for teaching vocabulary* (pp. 91–113). Routledge.

Ünaldı, İ, & Yüce, E. (2021). The relationship among vocabulary size, grammar proficiency, and critical thinking skills of adult language learners. *Adult Learning, 32*(2), 70–78. https://doi.org/10.1177/1045159520959473

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Snellings, P., Simis, A., & Stevenson, M. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: A structural equation modeling approach. *International Journal of Bilingualism, 7*(1), 7–25. https://doi.org/10.1177/13670069030070010201

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology, 96*(1), 19–30. https://doi.org/10.1037/0022-0663.96.1.19

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

Yang, W. (2011). Response validation of the reading comprehension subtest of the TEM 4 [Chinese]. *Foreign Language Teaching, 32*(06), 53–56.

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. (2017). In search of optimal cognitive diagnostic model(s) for ESL grammar test data. *Applied Measurement in Education, 30*(2), 82–101. https://doi.org/10.1080/08957347.2017.1283314

Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. *The Modern Language Journal, 96*(4), 558–575. https://doi.org/10.1111/j.1540-4781.2012.01398.x

Zhang, D., & Koda, K. (2013). Morphological awareness and reading comprehension in a foreign language: A study of young Chinese EFL learners. *System, 41*(4), 901–913. https://doi.org/10.1016/j.system.2013.09.009

Huilin Chen and Huan Mei Wish to submit an original research article entitled "How vocabulary knowledge and grammar knowledge influence L2 reading comprehension: A finer-grained perspective" for consideration by *European Journal of Psychology of Education*. We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

Current themes of research:

Application of cognitive diagnosis to language test data. Exploring relationships among vocabulary, grammar, and reading.

Most relevant publications in the field of Psychology of Education:

Chen, H., Cai, Y., & de la Torre, J. (2023). Investigating second language (L2) reading subskill associations: a cognitive diagnosis approach. *Language Assessment Quarterly, 20*(2), 166-189.

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.

Yi, Y. S. (2017). In search of optimal cognitive diagnostic model(s) for ESL grammar test data. *Applied Measurement in Education, 30*(2), 82-101. https://doi.org/10.1080/08957347.2017.1283314.

Raeisi-Vanani, A., & Baleghizadeh, S. (2022). The contributory role of grammar vs. vocabulary in L2 reading: An SEM approach. *Foreign Language Annals, 55*(2), 559-585. https://doi.org/10.1111/flan.12606.

Ünaldı, İ., & Yüce, E. (2021). The relationship among vocabulary size, grammar proficiency, and critical thinking skills of adult language learners. *Adult Learning, 32*(2), 70-78. https://doi.org/10.1177/1045159520959473.

Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: a structural equation modeling study. *The Modern Language Journal, 96*(4), 558-575. https://doi.org/10.1111/j.1540-4781.2012.01398.x.