Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Multidimensional Chinese Reading Ability Scale for Non-Chinese-Speaking Minority Students: Development and Validation

SIYI ZHU

s1147049@s.eduhk.hk

EdUHK: The Education University of Hong Kong https://orcid.org/0009-0000-8193-3192

Research Article

Keywords: Chinese reading ability; Scale development and validation; Non-Chinese-speaking minority students; Chinese as a second language; Measurement invariance

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11563694/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Multidimensional Chinese Reading Ability Scale for Non-Chinese-Speaking Minority Students: Development and Validation

Augustine Zhu SiYi[a]*

[a] *Department of Curriculum and Instruction, Education University of Hong Kong, Hong Kong, s1147049@s.eduhk.hk*

A Multidimensional Chinese Reading Ability Scale for Non-Chinese-Speaking Minority Students: Development and Validation

Abstract

To address the lack of multidimensional, psychometrically validated Chinese reading assessments for non-Chinese-speaking minority learners in China, this study developed and validated a 12-item Chinese Reading Ability Scale. Data came from 360 primary-school students (180 boys; 180 per school) in 10 classes from two ethnic-minority schools in Xinjiang. Analyses included parallel analysis, EFA, CFA, reliability, convergent/discriminant validity, measurement invariance across gender and school, common method bias, DIF, PCA-based unidimensionality/local independence, and Rasch item fit. Results showed that parallel analysis and EFA supported four factors (Items 1–3, 4–6, 7–9, 10–12), with excellent fit, loadings .811–.866, 71.1% variance, and low factor correlations (.080–.169). CFA confirmed the structure ($\chi^2(48)=29.385$, CFI=1.000, TLI=1.011, RMSEA=.000, SRMR=.019; loadings .808–.873). AVE .690–.726, CR .870–.888, HTMT .083–.175, and Fornell–Larcker supported validity; subscale α/ω .869–.888. Invariance was strict across gender and school. No gender DIF; no overall school DIF (Item 3 ETS B but nonsignificant). Common method bias was absent. PCA did not support strict unidimensionality (eigenvalue ratio 1.460), and local dependence was confined within clusters (max |Q3|=.691); all Rasch Infit/Outfit MNSQ .884–.996, ZSTD<2, PT-measure .360–.437. Theoretically, this study extends L2 Chinese reading models by confirming a multidimensional, measurement-invariant structure and challenging strictly unidimensional proficiency assumptions for an underrepresented minority population. Practically, the validated 12-item scale offers a population-sensitive diagnostic tool for identifying reading strengths and difficulties, enabling fair gender/school comparisons and informing differentiated assessment and instruction.

Keywords: Chinese reading ability; Scale development and validation; Non-Chinese-speaking minority students; Chinese as a second language; Measurement invariance

Introduction

Chinese reading assessment for non-Chinese-speaking minority learners in China remains insufficiently differentiated from assessment frameworks developed for broader Chinese-language populations. Although the Chinese Proficiency Test for Minorities (MHK) was established to provide a standardized system responsive to minority learners' acquisition of Chinese as an additional language, it primarily assesses overall proficiency rather than the multidimensional development of children's Chinese reading ability (Ministry of Education of the People's Republic of China, 2002; Peng et al., 2021). This issue is consequential because minority learners' Chinese literacy development is shaped by distinctive linguistic and instructional conditions (Zhang et al., 2020), while research with ethnic-minority CSL learners demonstrates that decoding, orthographic awareness, vocabulary, and higher-level comprehension make differentiated contributions to Chinese reading (Wong, 2019,

2020; Zhou, 2022, 2023).

Nevertheless, previous assessment research has predominantly examined international or additional-language learners outside mainland China, developing diagnostic measures and proficiency scales around component skills such as character recognition, word segmentation, vocabulary, comprehension, and reading strategies (Yang, 2021; Lin, 2023, 2024a, 2024b; Chen & Zhao, 2022). Existing research therefore provides a strong conceptual foundation but offers limited evidence for whether these constructs adequately represent non-Chinese-speaking minority children in China, particularly through a developmentally appropriate and psychometrically validated reading-specific framework.

Accordingly, this study aims to develop and validate a multidimensional Chinese Reading Ability Scale for non-Chinese-speaking minority students in China. Theoretically, the scale may extend current models of L2 Chinese reading by clarifying the dimensional structure of reading ability in an underrepresented sociolinguistic population. Practically, it may provide a population-sensitive tool for identifying reading strengths and difficulties and informing more appropriate assessment and instructional decisions.

Literature Review

Assessment Needs of Minority Chinese Reading Learners in China

Chinese reading development among non-Chinese-speaking minority learners in China cannot necessarily be evaluated through frameworks developed for native Chinese-speaking students. This concern has been recognized institutionally: the Chinese Proficiency Test for Minorities (MHK) was developed specifically to establish a scientific assessment system suited to minority learners' acquisition of Chinese as an additional language, rather than simply extending mainstream Chinese-language assessment to this population (Ministry of Education of the People's Republic of China, 2002; Peng et al., 2021). MHK subsequently became a national standardized assessment covering listening, reading, writing, and speaking, but its primary purpose is to measure overall proficiency in the national common language (National Education Examinations Authority, 2024; Peng et al., 2021) rather than to provide a child-focused, reading-specific account of Chinese literacy development.

Empirical evidence further indicates that minority learners' Chinese reading development is shaped by their distinctive linguistic and educational conditions. Using three waves of data from 3,869 ethnic-minority students in China, Zhang et al. (2020) found that reading–writing development differed across minority-language–Chinese and Chinese-only teaching models. Research with ethnic-minority CSL learners likewise demonstrates that orthographic awareness and character-decoding competence are important determinants of Chinese reading development (Wong, 2019, 2020; Wong et al., 2026). Evidence from Zhuang adolescents further shows substantial variation in Chinese reading literacy within a minority population (Li et al., 2020).

Thus, the issue is not the absence of assessment, but the limited availability of a **developmentally appropriate, reading-specific, and psychometrically validated framework** designed for non-Chinese-speaking minority children in China.

Chinese Reading Assessment for L2 learners

Research on Chinese reading assessment for second- and additional-language learners has developed from studies of component reading processes toward multidimensional proficiency scales and diagnostic instruments. Early empirical work identified character-reading fluency and word segmentation as important components of L2 Chinese reading comprehension (Shen & Jiang, 2013), while Kim and Kim (2014) examined comprehension across different meaning types. Subsequent diagnostic research developed and evaluated instruments targeting word reading, word segmentation, and text reading among low-level L2 Chinese learners (Yang, 2018, 2021a, 2021b). These studies established a stronger empirical basis for assessing Chinese-specific reading processes rather than treating reading as a unitary skill. Parallel work has addressed text difficulty through readability modeling (Sung et al., 2015), empirically validated proficiency hierarchies (Clifford & Cox, 2013; Lin, 2023a), and Chinese-specific graphemic competence (Guder, 2014). Recent assessment resources have further expanded from diagnostic measures to broader Chinese reading development, instruction, and assessment (Li & Zhang, 2022), proficiency testing (Cai et al., 2025), and graded reading-text characteristics (Yang, 2025).

Scale-development research, however, remains relatively concentrated in international CSL contexts. Lin (2020, 2024a) developed and validated a multidimensional Chinese L2 reading proficiency scale, with the 2024 study explicitly addressing the applicability of CSL reading assessment in US higher education. Lin (2023b) subsequently examined the dimensionality of Chinese L2 reading, identifying correlated lower-level decoding, textbase construction, and situation-model building components, while Lin (2024b) developed a measurement-driven scale specifically for US college-level CSL courses. Neubauer's (2020) review likewise characterizes beginning-level K–12 L2 Chinese reading research around L1-English learners, and Cai (2025) examines Chinese reading assessment in the Japanese learner context. Thus, the dominant assessment tradition has been constructed around learners acquiring Chinese as an additional or foreign language in international educational settings rather than around linguistically diverse non-Chinese-speaking children within China.

This distinction identifies an important population gap rather than an absence of Chinese-as-a-second-language research in China. Existing Chinese reading assessment frameworks and instruments provide substantial foundations for describing L2 reading proficiency, but they do not adequately establish whether these constructs and descriptors capture children who acquire Chinese while maintaining a different home language within China. A population-specific scale for non-Chinese-speaking minority students is therefore warranted to examine the applicability, dimensionality, and measurement validity of Chinese reading constructs in this

underrepresented sociolinguistic context. Table 1 presents the items developed for the Chinese Reading Ability Scale for Non-Chinese-Speaking Minority Students and the corresponding literature supporting each item.

Table 1. Items of the Chinese Reading Ability Scale for Non-Chinese-Speaking Minority Students and Corresponding Literature Support

Item Number	Item	Citation
D1. Chinese Orthographic and Word Recognition		
1	I can identify unfamiliar Chinese characters by using their components and structural features.	Wong (2020); Zhou (2023); Chen and Zhao (2022)
2	I can accurately read common Chinese words consisting of two or more characters.	Wong (2020); Liao et al. (2022); Yang (2021)
3	When reading Chinese sentences, I can correctly identify which characters should be grouped together as words.	Hao et al. (2022); Yang (2021); Chen (2021)
D2. Lexical and Syntactic Integration		
4	I can use the context to understand the meaning of unfamiliar Chinese words.	Zhang and Yang (2016); Chen and Zhang (2023); Yang et al. (2023)
5	I can use the meanings and components of Chinese characters to understand the meaning of unfamiliar words.	Zhou (2022); Yang et al. (2023); Chen and Zhao (2022)
6	I can use Chinese word order and connecting words to understand relationships between ideas in a sentence.	Zhou (2022); Yang et al. (2023); Lee-Thompson (2008)
D3. Text-Level Comprehension and Inference		
7	I can identify the main idea of a Chinese text.	Hao et al. (2022); Yang et al. (2023); Lin (2024)
8	I can connect information across different sentences to understand the overall meaning of a Chinese text.	Hao et al. (2022); Yang et al. (2023); Lin et al. (2024)

Item Number	Item	Citation
9	I can infer meanings that are not explicitly stated by using information provided in a Chinese text.	Yang et al. (2023); Hao et al. (2022); Lin (2024)
D4. Reading Monitoring and Strategic Processing		
10	When reading a Chinese text, I can monitor whether I have understood its content.	Yang et al. (2023); Zhao et al. (2021); Wang et al. (2023)
11	When I encounter Chinese content that I do not understand, I can use appropriate reading strategies to help myself understand it.	Lin et al. (2024); Lee-Thompson (2008); Wang et al. (2023)
12	After reading a Chinese text, I can identify and reorganize important information to check whether I have understood it.	Lin et al. (2024); Hao et al. (2022); Yang et al. (2023)

Figure 1. Measurement Validation Framework for the Present Study

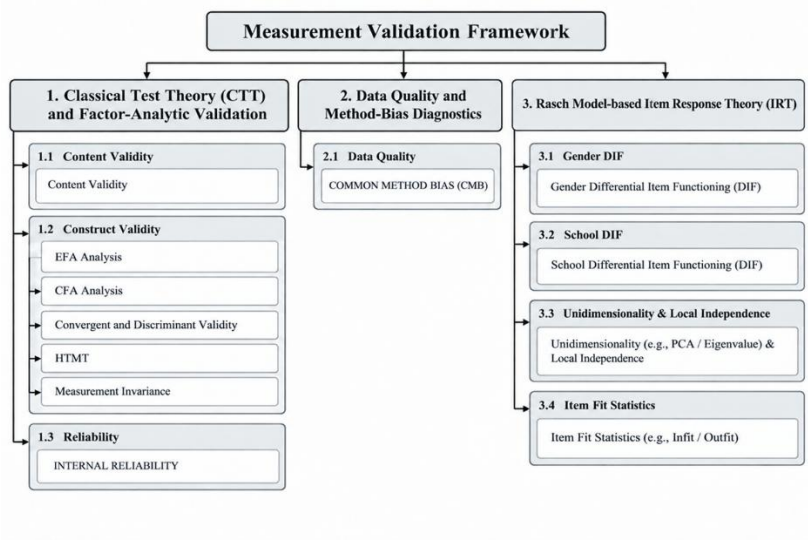


Figure 1 presents the measurement validation framework adopted in the present study, covering classical test theory, data-quality diagnostics, and Rasch-based analyses to establish the validity, reliability, and measurement quality of the instrument.

Research Sampling

After obtaining informed consent from the participating students, ethical approval was also obtained from the university to ensure that the study was conducted transparently and in accordance with relevant ethical requirements. A total of 360

primary-school students were recruited from two ethnic-minority primary schools located in two counties of the Xinjiang Uyghur Autonomous Region, China—five intact classes (n = 180) from a school in County A and five intact classes (n = 180) from a school in County B. Each of the ten participating classes contributed 36 students (18 boys and 18 girls), yielding a balanced design with respect to both school (A vs. B) and gender (male vs. female). The final analytic sample therefore consisted of 360 students (180 boys and 180 girls; 180 students from each of the two schools) nested within ten intact classrooms. No cases were excluded, and the 12-item instrument was administered in full to every participant.

Before student administration, three experts with relevant academic and professional backgrounds independently reviewed the proposed items for content validity. Following this expert review, all participating students were asked to read the same designated Chinese reading passage and complete the newly developed scale based on their own understanding and reading experience. The completed questionnaire data were subsequently collected and subjected to the measurement procedures presented in Figure 1.

Research Findings

Table 2. Expert Panel Profiles

Expert	Academic Qualification	Professional Position	Experience
Expert 1	PhD in Chinese Language/Curriculum Education	Full Professor	More than 20 years of university teaching experience
Expert 2	PhD in Chinese Language/Curriculum Education	Full Professor	More than 20 years of university teaching experience
Expert 3	PhD in Chinese Language/Curriculum Education	Education-sector practitioner	Professional experience in the education/training sector

Table 3. I-CVI + Modified Kappa

Item	Expert 1	Expert 2	Expert 3	Relevant Experts	I-CVI	Modified Kappa
1	4	4	4	3/3	1.00	1.00
2	4	3	4	3/3	1.00	1.00
3	4	4	4	3/3	1.00	1.00
4	4	4	4	3/3	1.00	1.00
5	4	4	4	3/3	1.00	1.00
6	4	4	4	3/3	1.00	1.00
7	4	4	4	3/3	1.00	1.00

Item	Expert 1	Expert 2	Expert 3	Relevant Experts	I-CVI	Modified Kappa
8	3	3	3	3/3	1.00	1.00
9	3	4	4	3/3	1.00	1.00
10	4	4	4	3/3	1.00	1.00
11	4	4	4	3/3	1.00	1.00
12	4	4	4	3/3	1.00	1.00

Note. rating scale: 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant.

Table 4. CVR

Item	Expert 1	Expert 2	Expert 3	Essential Experts (nen_e)	CVR
1	3	2	3	2/3	**0.33**
2	3	3	3	3/3	**1.00**
3	3	3	3	3/3	**1.00**
4	3	3	3	3/3	**1.00**
5	3	3	3	3/3	**1.00**
6	3	3	3	3/3	**1.00**
7	3	3	3	3/3	**1.00**
8	2	3	3	2/3	**0.33**
9	2	3	3	2/3	**0.33**
10	3	3	3	3/3	**1.00**
11	3	3	3	3/3	**1.00**
12	3	3	3	3/3	**1.00**

Note. rating scale: 1 = not essential, 2 = useful but not essential, 3 = essential.

Overall, the content validity results demonstrated strong expert support for the proposed scale. All 12 items achieved an I-CVI and Modified Kappa of 1.00, with S-CVI/Ave and S-CVI/UA also equal to 1.00. The mean CVR was 0.83, indicating generally strong agreement regarding item essentiality.

Table 5. Parallel Analysis for Determining the Number of Factors

Component	Observed solution
Number of factors suggested	**4**

Note. Parallel analysis was conducted using principal-axis factor analysis with 1,000 random iterations. The analysis suggested a four-factor solution.

As shown in Table 5, parallel analysis supported a four-factor solution, providing empirical evidence for retaining four factors.

Table 6. Exploratory Factor Analysis of the 12-Item Scale

Panel A. Model Fit Across Factor Solutions

Factor solution	χ^2	df	p	CFI	TLI	RMSEA	SRMR	BIC
1-factor	1698.264	54	< .001	0.273	0.110	0.2908	0.2548	1380.4
2-factor	1048.919	43	< .001	0.555	0.315	0.2549	0.1979	795.8
3-factor	536.170	33	< .001	0.777	0.552	0.2058	0.1377	341.9
4-factor	**16.713**	**24**	**.8606**	**1.000**	**1.009**	**0.0000**	**0.0076**	**−124.6**
5-factor	5.213	16	.9946	1.000	1.020	0.0000	0.0044	−89.0
6-factor	2.476	9	.9815	1.000	1.021	0.0000	0.0031	−50.5
7-factor	0.557	3	.9063	1.000	1.024	0.0000	0.0014	−17.1

Note. The EFA used maximum likelihood estimation with oblique GEOMIN rotation. Although solutions with five or more factors also showed very high fit, the four-factor solution was retained because it was the solution supported by parallel analysis and yielded a clear, interpretable simple structure. The four-factor model had $\chi^2(24) = 16.713$, $p = .861$, CFI = 1.000, TLI = 1.009, RMSEA = .0000 (90% CI [0.0000, 0.0235]), and SRMR = .0076.

Panel B. Standardized Factor Loadings and Factor Composition

Item	Factor 1	Factor 2	Factor 3	Factor 4
Item 1	**.847**	—	—	—
Item 2	**.866**	—	—	—
Item 3	**.842**	—	—	—
Item 4	—	**.849**	—	—
Item 5	—	**.846**	—	—
Item 6	—	**.854**	—	—
Item 7	—	—	**.813**	—
Item 8	—	—	**.845**	—
Item 9	—	—	**.858**	—
Item 10	—	—	—	**.811**
Item 11	—	—	—	**.814**
Item 12	—	—	—	**.864**

Note. Loadings shown are standardized loadings exceeding |.30|. The four-factor solution demonstrated a clear simple structure: Factor 1 comprised Items 1–3; Factor 2 comprised Items 4–6; Factor 3

comprised Items 7–9; and Factor 4 comprised Items 10–12. No salient cross-loadings above .30 were observed.

Panel C. Factor Correlations and Item-Level Commonality

Factor	F1	F2	F3	F4
F1	1.000	.080	.169	.141
F2	.080	1.000	.168	.121
F3	.169	.168	1.000	.130
F4	.141	.121	.130	1.000

Item	h^2	u^2
Item 1	.712	.288
Item 2	.758	.242
Item 3	.711	.289
Item 4	.722	.278
Item 5	.721	.279
Item 6	.729	.271
Item 7	.674	.326
Item 8	.701	.299
Item 9	.742	.258
Item 10	.674	.326
Item 11	.656	.344
Item 12	.744	.256

Note. The four factors were positively correlated, with correlations ranging from .080 to .169. Item communalities ranged from .656 to .758, whereas uniquenesses ranged from .242 to .344, indicating that the retained factors accounted for substantial proportions of item variance.

Panel D. Variance Explained by the Four Factors

Factor	SS loading	Proportion of variance	Cumulative variance
Factor 1	2.180	.182	.182
Factor 2	2.170	.181	.363
Factor 3	2.112	.176	.539
Factor 4	2.069	.172	.711

Note. The four factors accounted for a cumulative 71.1% of the variance, with individual proportions ranging from 17.2% to 18.2%.

Panel E. Residual Correlations

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12
Item 1	.29											
Item 2	.00	.24										
Item 3	.00	.00	.29									
Item 4	.00	.01	−.01	.28								
Item 5	.00	.00	.01	.00	.28							
Item 6	.01	−.01	.00	.00	.00	.27						
Item 7	.00	.00	.00	−.01	.00	.00	.33					
Item 8	−.01	.01	.00	.03	.00	−.02	.00	.30				
Item 9	.00	−.01	.00	−.02	.00	.02	.00	.00	.26			
Item 10	.01	−.01	.00	.00	.00	.00	−.01	.00	.01	.33		
Item 11	−.01	.00	.00	.00	.01	−.01	−.01	.00	.01	.00	.34	
Item 12	.00	.00	.00	.00	−.01	.00	.01	.00	−.01	.00	.00	.26

Note. Residual correlations were uniformly small, ranging from −.02 to .03 for off-diagonal elements, supporting the adequacy of the retained factor structure without evident substantial residual dependence.

Parallel analysis based on 1,000 random iterations supported a four-factor solution. Accordingly, maximum likelihood EFA with oblique GEOMIN rotation was conducted. The four-factor model demonstrated excellent fit, $\chi^2(24) = 16.713$, $p = .861$, CFI = 1.000, TLI = 1.009, RMSEA = .000 (90% CI [0.000, 0.024]), and SRMR = .008. The solution showed a clear simple structure, with Items 1–3 loading on Factor 1, Items 4–6 on Factor 2, Items 7–9 on Factor 3, and Items 10–12 on Factor 4. Standardized loadings ranged from .811 to .866, and the four factors accounted for 71.1% of the total variance. Factor correlations were relatively small (.080–.169), indicating that the four dimensions were empirically distinguishable.

Table 7. Confirmatory Factor Analysis of the Four-Factor Measurement Model

Panel A. Standardized Factor Loadings

Factor	Item	Standardized Loading	SE	*z*	*p*
DIM1	ITEM1	.839	.021	39.741	< .001
	ITEM2	.873	.021	41.190	< .001
	ITEM3	.843	.020	41.334	< .001
DIM2	ITEM4	.850	.021	41.354	< .001
	ITEM5	.848	.021	39.761	< .001
	ITEM6	.853	.020	43.247	< .001
DIM3	ITEM7	.822	.023	35.899	< .001
	ITEM8	.833	.023	36.852	< .001
	ITEM9	.863	.022	39.964	< .001
DIM4	ITEM10	.822	.025	33.330	< .001
	ITEM11	.808	.026	31.278	< .001
	ITEM12	.861	.022	38.468	< .001

Note. Standardized factor loadings are reported. All loadings were statistically significant ($p < .001$), ranging from **.808 to .873**, indicating strong relations between the observed items and their hypothesized latent dimensions.

Panel B. Standardized Factor Correlations

	DIM1	DIM2	DIM3	DIM4
DIM1	1.000	.085	.176	.148
DIM2	.085	1.000	.173	.125
DIM3	.176	.173	1.000	.135
DIM4	.148	.125	.135	1.000

Note. Factor correlations ranged from .085 to .176. DIM1–DIM2 ($p = .173$) and DIM2–DIM4 ($p = .050$) were not statistically significant, whereas the remaining factor correlations were statistically significant ($ps = .002$–$.025$). The relatively low correlations indicate that the four dimensions remain empirically distinguishable.

Panel C. Model Fit

χ^2	df	*p*	CFI	TLI	RMSEA	90% CI for RMSEA	SRMR	AIC	BIC	BIC2
29.385	48	.984	1.000	1.011	.000	[.000, .000]	.019	11061.576	11178.160	11082.984

Note. Fit indices are based on the scaled/robust estimates reported in the CFA output. The model demonstrated excellent fit, $\chi^2(48) = 29.385$, $p = .984$, CFI = 1.000, TLI = 1.011, RMSEA = .000, and SRMR = .019. The reported 90% confidence limits for RMSEA were both .000.

Panel D. Standardized Residual Covariances

	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6	ITEM7	ITEM8	ITEM9	ITEM10	ITEM11	ITE
ITEM1	.000											
ITEM2	.000	.000										
ITEM3	.002	−.001	.000									
ITEM4	−.043	.025	−.014	.000								
ITEM5	−.011	.045	.043	.000	.000							
ITEM6	−.046	−.004	−.013	.001	.000	.000						
ITEM7	−.017	.018	.015	.005	.030	.024	.000					
ITEM8	−.041	.002	−.003	−.005	−.021	−.049	.001	.000				
ITEM9	−.014	.009	.018	−.020	.013	.024	−.003	.002	.000			
ITEM10	.044	.028	.014	.022	.008	.013	.022	−.005	.033	.000		
ITEM11	−.035	−.024	−.039	.005	−.002	−.008	.000	−.023	.010	−.001	.000	
ITEM12	.004	.007	−.012	.005	−.029	−.006	.012	−.029	−.015	−.001	.002	.000

Note. Values are the standardized residual covariance matrix obtained using the Bentler residual specification. Off-diagonal residuals were small, ranging from **−.049 to .045**, providing no indication of substantial localized residual misfit.

Panel E. Modification Indices

Rank	Residual covariance	MI	EPC	SEPC.lv	SEPC.all	SEPC.nox
1	ITEM6 ~~ ITEM8	6.149	−.067	−.067	−.183	−.183
2	ITEM4 ~~ ITEM8	4.991	.059	.059	.165	.165
3	ITEM6 ~~ ITEM9	3.878	.052	.052	.154	.154
4	ITEM4 ~~ ITEM9	3.688	−.050	−.050	−.150	−.150

Note. The table reports the four largest modification indices among the factor-loading and covariance parameters examined. The largest modification index was 6.149 for the residual covariance between

ITEM6 and ITEM8. No modification was incorporated into the reported CFA model on the basis of these indices.

Panel F. Explained and Unexplained Item Variance

Factor	Item	R^2 (h^2)	Uniqueness (u^2)
DIM1	ITEM1	.704	.296
	ITEM2	.761	.239
	ITEM3	.711	.289
DIM2	ITEM4	.723	.277
	ITEM5	.719	.281
	ITEM6	.727	.273
DIM3	ITEM7	.676	.324
	ITEM8	.693	.307
	ITEM9	.744	.256
DIM4	ITEM10	.676	.324
	ITEM11	.652	.348
	ITEM12	.741	.259

Note. R^2 values represent the proportion of item variance explained by the corresponding latent factor, with values ranging from .652 to .761. Item uniquenesses ranged from .239 to .348.

A four-factor CFA was conducted to evaluate the hypothesized measurement structure. The model demonstrated excellent fit, $\chi^2(48) = 29.385$, $p = .984$, CFI = 1.000, TLI = 1.011, RMSEA = .000, and SRMR = .019. All 12 standardized factor loadings were statistically significant ($ps < .001$) and ranged from .808 to .873. The four factors showed relatively low correlations (.085–.176), supporting their empirical distinctiveness. Item-level R^2 values ranged from .652 to .761, indicating that the latent factors explained substantial proportions of variance in the observed indicators. Overall, the CFA supported the proposed four-factor structure comprising Items 1–3, Items 4–6, Items 7–9, and Items 10–12, respectively.

Table 8.
Panel A. Discriminant Validity of the Four-Factor Measurement Model

Factor	DIM1	DIM2	DIM3	DIM4
DIM1	**0.852**	0.085	0.176	0.148
DIM2	0.085	**0.850**	0.173	0.125
DIM3	0.176	0.173	**0.839**	0.135
DIM4	0.148	0.125	0.135	**0.831**

Note. Diagonal values (bold) represent the square roots of AVE, whereas off-diagonal values represent the standardized correlations between latent factors. The Fornell–Larcker criterion was satisfied for all four factors because each square root of AVE exceeded the corresponding inter-factor correlations.

Panel B. HTMT ratios

Factor Pair	HTMT
DIM1–DIM2	0.083
DIM1–DIM3	0.175
DIM1–DIM4	0.146
DIM2–DIM3	0.173
DIM2–DIM4	0.126
DIM3–DIM4	0.136

Note. HTMT = heterotrait–monotrait ratio. All HTMT values were below .85, satisfying the strict criterion used in the analysis. All values were also below .90.

Table 9. Latent Factor Correlations and 95% Confidence Intervals

Factor Pair	*r*	95% CI	Contains 1
DIM1–DIM2	0.085	[−0.037, 0.208]	No
DIM1–DIM3	0.176	[0.060, 0.293]	No
DIM1–DIM4	0.148	[0.032, 0.263]	No
DIM2–DIM3	0.173	[0.062, 0.283]	No
DIM2–DIM4	0.125	[0.000, 0.249]	No
DIM3–DIM4	0.135	[0.017, 0.254]	No

Note. r represents the standardized latent factor correlation. The 95% confidence intervals are based on the standardized solution. None of the confidence intervals for the six pairs of distinct factors contained 1.

The four factors demonstrated satisfactory convergent validity. Based on table 8, AVE values ranged from .690 to .726, exceeding the .50 criterion, while CR values ranged from .870 to .888, exceeding the .70 criterion. Thus, all four dimensions met the specified criteria for convergent validity. Discriminant validity was also supported. The square roots of AVE ranged from .831 to .852 and exceeded the corresponding inter-factor correlations (.085–.176), satisfying the Fornell–Larcker criterion. The HTMT ratios ranged from .083 to .175, with all values below the strict .85 criterion. In addition, table 9 suggested that the 95% confidence intervals for all six correlations between distinct factors did not contain 1. Taken together, these results support the distinctiveness of the four dimensions in the measurement model.

Table 10. Measurement Invariance Across Gender

Invariance Model	χ^2	df	*p*	CFI	TLI	RMSEA	SRMR
Configural	85.906	96	.760	1.000	1.006	.000	.033
Metric	102.971	104	.510	1.000	1.001	.000	.039
Scalar	108.813	112	.568	1.000	1.002	.000	.039
Strict	128.622	124	.370	.998	.998	.014	.036

Note. Configural, metric, scalar, and strict invariance models all converged successfully. Changes in fit indices were evaluated across successive levels of invariance. All three comparisons met the criteria specified in the analysis and were classified as *PASS*.

Table 11. Changes in Fit Indices Across Gender Invariance Levels

Comparison	ΔCFI	ΔRMSEA	ΔSRMR	Result
Metric vs. Configural	0.000	0.000	0.006	PASS
Scalar vs. Metric	0.000	0.000	0.000	PASS
Strict vs. Scalar	−0.002	0.014	−0.003	PASS

Note. ΔCFI, ΔRMSEA, and ΔSRMR represent changes in the corresponding fit indices between successive invariance models. All comparisons were classified as *PASS* according to the criteria specified in the analysis.

Table 12. Measurement Invariance Across Schools

Invariance Model	χ^2	df	*p*	CFI	TLI	RMSEA	SRMR
Configural	77.506	96	.917	1.000	1.011	.000	.030
Metric	84.076	104	.924	1.000	1.011	.000	.032
Scalar	89.566	112	.942	1.000	1.012	.000	.033
Strict	98.524	124	.956	1.000	1.012	.000	.034

Note. Configural, metric, scalar, and strict invariance models all converged successfully. Changes in fit indices were evaluated across successive levels of invariance. All three comparisons met the criteria specified in the analysis and were classified as *PASS*.

Table 13. Changes in Fit Indices Across School Invariance Levels

Comparison	ΔCFI	ΔRMSEA	ΔSRMR	Result
Metric vs. Configural	0.000	0.000	0.002	PASS
Scalar vs. Metric	0.000	0.000	0.001	PASS
Strict vs. Scalar	0.000	0.000	0.001	PASS

Note. ΔCFI, ΔRMSEA, and ΔSRMR represent changes in the corresponding fit indices between successive invariance models. All comparisons were classified as *PASS* according to the criteria specified in the analysis.

Table 14. Summary of Measurement Invariance Testing Across Gender and School

Grouping Variable	Invariance Level	ΔCFI	ΔRMSEA	ΔSRMR	Result
Gender	Metric	0.000	0.000	0.006	PASS
Gender	Scalar	0.000	0.000	0.000	PASS
Gender	Strict	−0.002	0.014	−0.003	PASS
School	Metric	0.000	0.000	0.002	PASS
School	Scalar	0.000	0.000	0.001	PASS
School	Strict	0.000	0.000	0.001	PASS

Note. ΔCFI, ΔRMSEA, and ΔSRMR denote changes in fit indices across successive levels of measurement invariance. All comparisons were classified as *PASS* according to the criteria specified in the analysis.

Measurement invariance was examined across gender and school at the configural, metric, scalar, and strict levels. Across gender, all four models showed satisfactory fit, with CFI values ranging from .998 to 1.000, TLI values from .998 to 1.006, RMSEA values from .000 to .014, and SRMR values from .033 to .039. The changes from configural to metric, metric to scalar, and scalar to strict invariance were all classified as PASS (ΔCFI = 0.000, 0.000, and −0.002, respectively).

Across schools, all four models also showed satisfactory fit, with CFI = 1.000, TLI ranging from 1.011 to 1.012, RMSEA = .000, and SRMR ranging from .030 to .034. The metric, scalar, and strict comparisons were all classified as PASS, with ΔCFI values of 0.000 across all three comparisons. Overall, the results supported configural, metric, scalar, and strict invariance across both gender and school.

Table 15. Internal Reliability of the Four-Factor Scale

Scale / Subscale	*k*	Cronbach's α	McDonald's ω	Standardized α	G6
DIM1	3	.888	.888	.888	.841
DIM2	3	.887	.887	.887	.839
DIM3	3	.877	.877	.877	.827
DIM4	3	.869	.870	.869	.817
Total scale	**12**	**.762**	**.967**	**.762**	**.879**

Note. *k* = number of items; α = Cronbach's alpha; ω = McDonald's omega; G6 = Guttman's G6 coefficient. Cronbach's α and McDonald's ω for the four subscales are from the primary internal reliability analysis. Standardized α and G6 are from the psych::alpha cross-check. For the total scale, α

and ω are from the primary total-scale analysis, whereas standardized α and G6 are from the psych::alpha cross-check. The analysis classified α ≥ .70 and ω ≥ .70 as meeting the specified reliability criterion; all four subscales met both criteria.

As shown in Table 12, the four subscales demonstrated consistent internal reliability across the reported coefficients. Cronbach's α ranged from .869 to .888, and McDonald's ω ranged from .869 to .888 across the four subscales. The 12-item total scale yielded a Cronbach's α of .762 and a McDonald's ω of .967. The psych::alpha cross-check produced standardized α values ranging from .869 to .888 across the four subscales and .762 for the total scale, with G6 coefficients ranging from .817 to .841 across the subscales and .879 for the total scale.

Table 16. Assessment of Common Method Bias

Method	Key Statistic	Observed Value	Criterion	Decision
Harman’s single-factor test (PCA)	Variance explained by first component	27.69% < 50%		PASS
One-factor vs. four-factor CFA	ΔCFI	+.748 > .05		PASS
	ΔBIC	−1637.1 < −10		PASS
Variance inflation factor (VIF)	Maximum VIF	2.743 < 3.3		PASS

Note. PCA = principal component analysis; CFA = confirmatory factor analysis; VIF = variance inflation factor; CFI = comparative fit index; BIC = Bayesian information criterion. The first principal component accounted for 27.69% of the total variance. The four-factor model showed substantially better fit than the one-factor model (ΔCFI = +.748; ΔBIC = −1637.1). All item-level VIF values were below 3.3, with a maximum of 2.743. Based on these three diagnostics, the analysis indicated no severe common method bias. The common latent factor analysis was not used as a primary diagnostic because the model produced estimation and identification warnings, including failure to invert the information matrix and a non-positive-definite latent covariance matrix.

Three complementary diagnostics were used to assess potential common method bias. Harman’s single-factor test showed that the first principal component accounted for 27.69% of the total variance, below the 50% criterion. In addition, the four-factor CFA substantially outperformed the one-factor model (ΔCFI = .748; ΔBIC = −1637.1). Finally, all item-level VIF values were below 3.3, with a maximum of 2.743. Collectively, these results provided no indication of severe common method bias.

Table 17. Differential Item Functioning by Gender

Item	MH χ^2	*p*	ETS	Wald DIF
ITEM1	0.162	.6873	A	No
ITEM2	0.000	.9995	A	No
ITEM3	0.000	.9875	A	No
ITEM4	0.234	.6288	A	No
ITEM5	1.026	.3112	A	No
ITEM6	0.002	.9630	A	No
ITEM7	1.584	.2082	A	No
ITEM8	0.003	.9588	A	No
ITEM9	0.012	.9132	A	No
ITEM10	0.136	.7126	A	No
ITEM11	2.254	.1333	A	No
ITEM12	1.822	.1771	A	No
Overall DIF (Andersen LR test)	**7.349**	**.9204**	—	**No**

Note. MH = Mantel–Haenszel; ETS = Educational Testing Service classification. ETS A indicates negligible DIF. The Wald test showed no significant item-level DIF; all reported threshold-parameter *p* values exceeded .05 (minimum $p = .142$). The overall LR test was nonsignificant.

Table 18. Differential Item Functioning by School

Item	MH χ^2	*p*	ETS	Wald DIF
ITEM1	0.788	.3747	A	No
ITEM2	1.930	.1648	A	No
ITEM3	2.509	.1132	B	No
ITEM4	0.010	.9200	A	No
ITEM5	0.366	.5454	A	No
ITEM6	0.285	.5933	A	No
ITEM7	0.003	.9576	A	No
ITEM8	0.849	.3568	A	No
ITEM9	0.052	.8199	A	No
ITEM10	0.730	.3928	A	No
ITEM11	0.006	.9372	A	No
ITEM12	0.563	.4532	A	No
Overall DIF (Andersen LR test)	**12.044**	**.6028**	—	**No**

Note. MH = Mantel–Haenszel; ETS = Educational Testing Service classification. ETS A indicates negligible DIF, whereas ETS B indicates moderate DIF. ITEM3 was classified as ETS B, although its

MH test was nonsignificant (p = .1132), and the Wald test also showed no significant item-level DIF. The overall LR test was nonsignificant.

No evidence of DIF was found by gender. The Andersen LR test was nonsignificant, $\chi^2(14) = 7.349$, $p = .9204$, and all 12 items were classified as ETS A by the MH procedure.No statistically significant overall DIF was detected, $\chi^2(14) = 12.044$, $p = .6028$. All Wald tests were nonsignificant. Although ITEM3 received an ETS B classification, its MH test was nonsignificant (p = .1132), so it does not provide statistically significant evidence of DIF.

Table 19. *Unidimensionality Assessment Based on Principal Component Analysis*

Criterion	Result	Decision
First eigenvalue	3.322	—
Second eigenvalue	2.275	—
First-to-second eigenvalue ratio	1.460	Fail
Variance explained by first component	27.69%	Pass
First-component loadings	0.254–0.326	Pass
All first-component loadings positive	Yes	Pass

Note. The first-to-second eigenvalue ratio was evaluated against a ratio of 3.00, whereas 20% was used as a descriptive benchmark for the variance explained by the first component. The ratio of 1.460 does not support a unidimensional interpretation. The PCA results should therefore be interpreted as evidence against strict unidimensionality rather than as a failure of item coherence.

Table 20. *Local Independence Assessment Using Residual Correlations and Item–Total Correlations*

Criterion	Result	Decision
Maximum |Q3|	0.691	Fail
Q3 pairs with |Q3| ≥ .20	12	Flagged
Minimum item–total correlation	0.360	Pass
All item–total correlations > .30	Yes	Pass

Note. Q3 denotes Yen's residual correlation statistic. The conventional screening criterion of |Q3| < .20 was used. The 12 flagged pairs were concentrated within the four hypothesized three-item clusters (Items 1–3, 4–6, 7–9, and 10–12). Item–total correlations ranged from .360 to .437.

PCA results did not support strict unidimensionality, with a first-to-second eigenvalue ratio of 1.460, although the first component explained 27.69% of the variance and all first-component loadings were positive. Local-independence diagnostics also identified elevated residual associations (maximum |Q3| = .691),

whereas all item–total correlations exceeded .30. Notably, the flagged Q3 associations were concentrated within the four hypothesized item clusters, suggesting that the residual dependence may reflect the underlying multidimensional structure rather than isolated item redundancy.

Table 21. Rasch Item Fit Statistics and Point-Measure Correlations

Item	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PT-Measure
ITEM1	0.988	−0.157	0.993	−0.085	0.369
ITEM2	0.936	−0.969	0.943	−0.847	0.426
ITEM3	0.919	−1.229	0.924	−1.158	0.401
ITEM4	0.896	−1.611	0.901	−1.513	0.379
ITEM5	0.904	−1.472	0.906	−1.434	0.394
ITEM6	0.948	−0.773	0.945	−0.827	0.369
ITEM7	0.887	−1.747	0.884	−1.791	0.433
ITEM8	0.937	−0.955	0.945	−0.825	0.386
ITEM9	0.912	−1.347	0.928	−1.090	0.437
ITEM10	0.970	−0.435	0.969	−0.449	0.414
ITEM11	0.957	−0.638	0.966	−0.494	0.360
ITEM12	0.994	−0.071	0.996	−0.044	0.384

Note. MNSQ = mean-square fit statistic; ZSTD = standardized fit statistic; PT-Measure = point-measure correlation. Infit and Outfit MNSQ values between 0.50 and 1.50 were considered acceptable. Absolute ZSTD values < 2.00 and PT-Measure correlations > .30 were used as supplementary criteria. All 12 items met all criteria; no item showed substantive misfit.

All 12 items demonstrated acceptable Rasch fit, with Infit MNSQ values ranging from 0.887 to 0.994 and Outfit MNSQ values from 0.884 to 0.996. All absolute ZSTD values were below 2, and PT-Measure correlations ranged from .360 to .437. Thus, no item exhibited substantive misfit.

Implications of the Study

Theoretically, this study advances Chinese second-language reading research by operationalizing reading ability as a multidimensional construct among non-Chinese-speaking minority students in China. The four-factor structure challenges strictly unidimensional interpretations of reading proficiency. However, evidence from two schools in Xinjiang cannot establish a universal model, and independent replication remains necessary.

Practically, the 12-item scale offers a concise tool for identifying potential reading-related strengths and difficulties and informing differentiated instruction. Further validation of criterion-related validity, predictive accuracy, and sensitivity to instructional change is necessary before high-stakes use. Its contribution is therefore best understood as an initial, population-sensitive assessment framework rather than a definitive measure of minority students' reading competence.

Limitations and Future Research

This study has two limitations. First, EFA and CFA were conducted using the same sample of 360 students, potentially inflating the apparent stability of the factor structure due to sample-specific characteristics. Future research should validate the proposed model using independent samples and recruit students from diverse ethnic, regional, and educational backgrounds to establish its generalizability.

Second, paper-based administration and manual data entry were time-consuming and potentially susceptible to transcription errors. Although data screening was conducted, such errors cannot be entirely excluded without independent verification. Future studies could adopt secure digital administration or double-entry procedures to improve efficiency and accuracy.

Ethics Declarations

Ethics Statement

Ethical approval for this study was granted by the ethics committee at Minnan University of Science and Technology. No separate ethics approval number was issued. Written informed consent was obtained from the participating students with corresponding guardian and head teacher consent. The study was conducted in accordance with the applicable institutional ethical requirements and relevant ethical principles for research involving human participants.

Consent for publication

Consent for publication of anonymized data was obtained from all participants and, where applicable, their guardians.

Declaration of generative AI use

Generative AI was used solely to assist with language editing and improve the clarity and readability of the manuscript. It was not used for study design, literature synthesis, methodology development, data collection, data analysis, or the generation of research findings. The author takes full responsibility for the accuracy, integrity, and originality of the manuscript.

Data availability statement

The dataset supporting the findings of this study has been deposited in Zenodo

(https://doi.org/10.5281/zenodo.23250452) under restricted access to protect participant confidentiality. Access will be granted where appropriate and in accordance with ethical and institutional requirements.

Competing interests

The author declares no competing interests.

Funding

This research received no external funding.

References

Chen, X., & Zhao, J. (2022). Reading-related skills associated with acquisition of Chinese as a second/foreign language: A meta-analysis. Frontiers in Psychology, 13, 783964. https://doi.org/10.3389/fpsyg.2022.783964

Chen, K. (2021). Can word segmentation and length of text affect L2 reading? Evidence from advanced L2 learners of Chinese and the three reading models. Theory and Practice in Language Studies, 11(9), 1109–1113. https://doi.org/10.17507/tpls.1109.18

Chen, T., & Zhang, D. (2023). Different aspects of vocabulary depth knowledge in L2 Chinese reading comprehension: Comparing higher- and lower-proficiency readers. Foreign Language Annals, 56, 786–806. https://doi.org/10.1111/flan.12687

Cai, J., Han, Y., & Jiang, X. (2025). ClozCHI: A cloze test for measuring L2 Chinese proficiency from novice to advanced levels. Behavior Research Methods, 57(12), 325. https://doi.org/10.3758/s13428-025-02834-9

Cai, Q. Y. (2025). Analyzing the alignment of Japanese Chinese Proficiency Test Level 3 reading items with HSK standards: A comprehensive vocabulary and thematic evaluation. Research Square. https://doi.org/10.21203/rs.3.rs-6850665/v1

Clifford, R., & Cox, T. L. (2013). Empirical validation of reading proficiency guidelines. Foreign Language Annals, 46(1), 45–61. https://doi.org/10.1111/flan.12033

Guder, A. (2014). Reading competence and graphemic competence: Impacts of the Chinese writing system on designing competence descriptors for the “European Benchmarks for the Chinese Language” (EBCL) project. CHUN–Chinesischunterricht, 29, 5–29. No DOI identified.

Hao, M., Fang, X., Sun, Z., & Liu, Y. (2022). The extended simple view of reading in adult learners of Chinese as a second language. Frontiers in Psychology, 13, 846967. https://doi.org/10.3389/fpsyg.2022.846967

Kim, M. S., & Kim, H. J. (2014). Investigating Chinese language learners' reading comprehension for different meaning types. GEMA Online Journal of Language Studies, 14(1), 77–100. https://doi.org/10.17576/GEMA-2014-1401-06

Lee-Thompson, L.-C. (2008). An investigation of reading strategies applied by American learners of Chinese as a foreign language. Foreign Language Annals, 41(4), 702–721. https://doi.org/10.1111/j.1944-9720.2008.tb03326.x

Li, L., Chen, X., Wu, D., & Li, H. (2020). Effects of attending preschool on adolescents' reading literacy: Evidence from the ethnic minority children in China. Children and Youth Services Review, 116, 105211. https://doi.org/10.1016/j.childyouth.2020.105211

Liao, X., Loh, E. K. Y., & Cai, M. (2022). Lexical orthographic knowledge mediates the relationship between character reading and reading comprehension among learners with Chinese as a second language. Frontiers in Psychology, 13, 779905. https://doi.org/10.3389/fpsyg.2022.779905

Lin, J. (2024). Developing a reading proficiency scale for US college-level CSL courses: A complement to the ACTFL guidelines. Language Teaching Research. https://doi.org/10.1177/13621688241266141

Lin, J., Gao, G., & Huang, T. (2024). Top-down Chinese as a second language reading strategies. Chinese as a Second Language Research, 13(1), 1–27. https://doi.org/10.1515/caslar-2024-0001

Li, L., & Zhang, D. (Eds.). (2022). Reading in Chinese as an additional language: Learners' development, instruction, and assessment. Routledge. https://doi.org/10.4324/9781003029038

Lin, J. (2020). Development and validation of a reading proficiency scale for Chinese as a second language (Doctoral dissertation, The University of North Carolina at Chapel Hill). No DOI identified. The dissertation is explicitly cited as an unpublished doctoral dissertation in Lin's later scale-development article.

Lin, J. (2023a). ACTFL Chinese reading proficiency guidelines: Verifying the difficulty hierarchy. Foreign Language Annals, 56(3), 667–689. https://doi.org/10.1111/flan.12680

Lin, J. (2023b). Assessing the dimensionality of Chinese as a second language reading: A confirmatory factor analysis approach. Chinese as a Second Language Research, 12(2), 173–204. https://doi.org/10.1515/caslar-2023-2001

Lin, J. (2024a). Developing a reading proficiency scale for Chinese as a second language: A confirmatory factor analysis approach. Language Awareness, 33(3), 597–624. https://doi.org/10.1080/09658416.2023.2291521

Lin, J. (2024b). Developing a reading proficiency scale for US college-level CSL courses: A complement to the ACTFL guidelines. Language Teaching Research. https://doi.org/10.1177/13621688241266141

Ministry of Education of the People's Republic of China. (2002). 在有关省市试行中国少数民族汉语水平等级考试的通知 [Notice on piloting the Chinese Proficiency Test for Minorities in relevant provinces and municipalities]. Ministry of Education of the People's Republic of China. Official MHK policy information

Neubauer, D. E. (2020). L2 Chinese reading comprehension among beginning-level, K–12 learners: A literature review. Chinese Language Teaching Methodology and Technology, 3(2), Article 4. No DOI identified.

Peng, Y., Yan, W., & Cheng, L. (2021). Hanyu Shuiping Kaoshi (HSK): A multi-level, multi-purpose proficiency test. Language Testing, 38(4), 511–530. https://doi.org/10.1177/0265532220957298

Shen, H. H., & Jiang, X. (2013). Character reading fluency, word segmentation accuracy, and reading comprehension in L2 Chinese. Reading in a Foreign Language, 25(1), 1–25. https://doi.org/10.64152/10125/66679

Sung, Y. T., Lin, W. C., Dyson, S. B., Chang, K. E., & Chen, Y. C. (2015). Leveling L2 texts through readability: Combining multilevel linguistic features with the CEFR. The Modern Language Journal, 99(2), 371–391. https://doi.org/10.1111/modl.12213

Wong, Y. K. (2019). Role of decoding competence in the Chinese reading comprehension development of ethnic minority students in Hong Kong. *International Journal of Bilingual Education and Bilingualism, 22*(8), 1016–1029. https://doi.org/10.1080/13670050.2017.1329273

Wong, Y. K. (2020). The relationship between orthographic awareness and Chinese reading in Hong Kong's young CSL learners. *ECNU Review of Education, 3*(4), 678–693. https://doi.org/10.1177/2096531120930485

Wong, Y. K., Zhou, Y., & Zhang, D. (2026). Role of metalinguistic awareness in young learners of Chinese as a second language in character reading and spelling: A one-year longitudinal study. *Reading and Writing*. https://doi.org/10.1007/s11145-026-10801-9

Wang, Y., Hu, J., An, Z., Li, C., & Zhao, Y. (2023). The influence of metacognition monitoring on L2 Chinese audiovisual reading comprehension. Frontiers in Psychology, 14, 1133003. https://doi.org/10.3389/fpsyg.2023.1133003

Wong, Y. K. (2019). Role of decoding competence in the Chinese reading comprehension development of ethnic minority students in Hong Kong. International Journal of Bilingual Education and Bilingualism, 22(8), 1016–1029. https://doi.org/10.1080/13670050.2017.1329273

Wong, Y. K. (2020). The relationship between orthographic awareness and Chinese reading in Hong Kong's young CSL learners. ECNU Review of Education, 3(4), 678–693. https://doi.org/10.1177/2096531120930485

Wong, Y. K., Zang, X., & Inoue, T. (2023). Promoting foundational linguistic skills for reading development in young Chinese language learners: A 1-year intervention study. Journal of Research in Reading, 46(3), 247–277. https://doi.org/10.1111/1467-9817.12423

Yang, L., Xiong, Y., & Chen, Q. (2023). The role of linguistic and cognitive skills in reading Chinese as a second language: A path analysis modeling approach. Frontiers in Psychology, 14, 1131913. https://doi.org/10.3389/fpsyg.2023.1131913

Yang, S. (2021). Diagnosing reading problems for low-level Chinese as second language learners. System, 97, 102433. https://doi.org/10.1016/j.system.2020.102433

Yang, S. (2018). Constructing informal diagnostic reading assessment instruments for lower-level Chinese as second language readers (Doctoral dissertation, University of Iowa). No DOI identified.

Yang, S. (2021a). Constructing diagnostic reading assessment instruments for low-level Chinese as second language learners. Reading in a Foreign Language, 33(2), 212–237. https://doi.org/10.64152/10125/67401

Yang, S. (2021b). Diagnosing reading problems for low-level Chinese as second language learners. System, 97, 102433. https://doi.org/10.1016/j.system.2020.102433

Yang, L. J. (2025). 《国际中文教育中文阅读分级标准》(青少年)读物语言特征体系研究. Journal of International Chinese Teaching, (2), 14.

Zhang, Y., Peng, H., & Bian, Y. (2020). Interplay between reading and writing under different teaching models: A study based on Chinese learning by China's ethnic minorities. *Frontiers in Psychology, 11*, 2150. https://doi.org/10.3389/fpsyg.2020.02150

Zhang, D., & Yang, X. (2016). Chinese L2 learners' depth of vocabulary knowledge and its role in reading comprehension. *Foreign Language Annals, 49*(4), 699–715. https://doi.org/10.1111/flan.12225

Zhao, A., Guo, Y., Sun, S., Lai, M. H. C., Breit, A., & Li, M. (2021). The contributions of language skills and comprehension monitoring to Chinese reading comprehension: A longitudinal investigation. *Frontiers in Psychology, 12*, 625555. https://doi.org/10.3389/fpsyg.2021.625555

Zhou, J. (2022). The contribution of morphological awareness and vocabulary knowledge to Chinese as a second language reading comprehension: A path analysis. *Journal of Psycholinguistic Research, 51*(1), 55–74. https://doi.org/10.1007/s10936-021-09810-2

Zhou, J. (2022). The effects of syntactic awareness to L2 Chinese passage-level reading comprehension. *Frontiers in Psychology, 12*, 783827. https://doi.org/10.3389/fpsyg.2021.783827

Zhou, J. (2023). The contribution of radical knowledge and character recognition to L2 Chinese reading comprehension. *Journal of Psycholinguistic Research, 52*(2), 445–475. https://doi.org/10.1007/s10936-022-09880-w