A Preliminary Research in Developing and Validation of a Framework for Assessing Linguistic Competencies

Suet-Ching Soon[1] and Chia-Ling Hsu [2,*]

[1] National United University, Taiwan
[2] National Academy for Educational Research, Taiwan
[1]ssoon@gm.nuu.edu.tw
[2,*]clhsu@mail.naer.edu.tw

Abstract. Syntactic knowledge is a fundamental element of linguistic competence and plays a vital role for individuals preparing to teach or work in language-focused professions. This study developed a 40-item instrument to measure syntactic knowledge in Chinese and examine its psychometric properties using the Rasch model with a sample of 112 undergraduate students. Each item was dichotomously scored (1 = correct, 0 = incorrect). The Rasch analysis yielded several key findings: (a) all 40 items effectively measured the intended construct of syntactic knowledge in Chinese sentential comprehension; (b) fit indices (unweighted mean square error and weighted mean square error) indicated good model-data fit, with values falling within the accepted range of 0.7-1.3; (c) Person Separation Reliability (PSR) demonstrated excellent reliability, exceeding .99; and (d) more than 95% of item difficulty values were below -1.0 logit. Collectively, these results validated the psychometric properties of the instrument with respect to both construct validity and reliability at the item level. The findings support the use of this tool for assessing students' syntactic knowledge in Chinese and highlight its potential as a diagnostic instrument in classroom settings, particularly for identifying learners with low to moderate ability levels and informing instruction adjustment or targeted interventions

Keywords: Rasch Analysis, instrument validation, Chinese syntactic knowledge.

1 Introduction

The demand for professional Chinese language teachers has increased significantly over the past couple of decades due to various reasons for the growing trend of learning Chinese. Accordingly, numerous studies have investigated learner competence from the perspective of second language acquisition, exploring areas such as language skills development, proficiency frameworks, pedagogical implications, and materials developments [1, 7, 8, 11, 16, 20, 21, 24, 40, 43]. A qualified Chinese language instructor must demonstrate not only expertise in pedagogical practices but also a deep understanding of linguistic knowledge, which is one of the fundamental requirements for effective language teaching [8, 22, 42]. Since language use embodies a speaker's

© The Author(s) 2026
C. Y. Lye et al. (eds.), *Proceedings of the Pacific Rim Objective Measurement Symposium (PROMS 2025)*, Atlantis Highlights in Social Sciences, Education and Humanities 56,
https://doi.org/10.2991/978-94-6239-834-4_11

knowledge of the underlying linguistic systems, syntax can be understood as the set of rules and structures governing the formation of grammatical sentences [13, 18]. An individual's language use therefore provides evidence of both their acquisition and their linguistic competence. At the same time, observable language performance reflects how this competence is realized in practice, thereby offering a window into the interaction between underlying grammatical knowledge and actual communicative behavior.

The present study aims to develop, evaluate, and validate an instrument for assessing Chinese linguistic competence through the application of Rasch model in construct development. The methodological framework involved several stages. The process began with drafting items to capture key dimensions of Chinese linguistic competence, which were then refined through expert review and pilot testing to ensure content relevance and clarity. Next, item calibration was conducted using Rasch model to estimate item parameters (i.e., difficulty), and precisely evaluate item functioning. The calibration allowed for an examination of construct validity at the item through model-fit indices, ensuring that the instrument accurately reflects the intended latent traits. Finally, person-separation reliability was estimated to assess the instrument's ability to differentiate among examinees with varying proficiency levels. Data were collected from a sample of undergraduate students majoring in Chinese – representing prospective Chinese language instructors, to ensure a relevant validation population. Collectively, these procedures yield a psychometrically instrument backed by strong evidence of both validity and reliability.

2 Linguistic Competence

Linguistic competence has received considerable scholarly attention for the theoretical insights it conveyed [15, 17, 32, 42]. In distinguishing competence from performance, prior studies have generally defined competence as the underlying ability to use language, whereas performance comprises the observable behaviors through which tasks are accomplished [23]. Meanwhile, from this perspective, competence can be viewed as "a bundle of cognitively controlled abilities or skills…[that] implies both knowledge and the ability and disposition to solve problem in that domain" [27].

To understand how language is used or acquired, Chomsky argued that researchers must isolate this underlying cognitive system, "… we must abstract for separate and independent study a cognitive system, a system of knowledge and belief, that develops in early childhood and that interacts with many other factors to determine the kinds of behavior that we observe; to introduce a technical term, we must isolate and study the system of linguistic competence that underlies behavior but that is not realized in any direct or simple way in behavior" [14]. Central to this framework is the concept of Universal Grammar (UG), which posits that all human languages share common structural principles, such as the presence of noun phrases, verb phrases and sentences [14]. Within this scope, linguistic competence is defined as the initialize system of syntactic rules that govern sentence formation. These rules encompass phenomena such

as word order, agreement, movement, subordination. For example, English is characterized as a Subject-Verb-Object (SVO) language, featuring strict subject-verb agreement (for example, singular subjects requiring singular verbs) and syntactic movement, as seen when transforming declarative sentences into interrogatives.

Although native speakers are often assumed to possess inherent competence in their first language, proficiency in additional languages varies considerably [27]. To objectively compare an individual's competence across native and non-native languages, an empirical framework is required; because internal competence cannot be observed directly, it must be inferred through measurable performance [39]. Consequently, assessing these competencies requires the development of cognitive models and psychometric instruments that can explicitly link theoretical constructs to assessment outcomes [23].

This theoretical approach stands in contrasted to general language proficiency frameworks, which focus primarily on a learners' acquisition of language skills and the establishment of standardized benchmarks. While functional proficiency frameworks facilitate the creation of reliable evaluation tools and instructional materials, they remain insufficient for assessing pre-service instructors or language professionals. Since these benchmarks are designed to measure functional communicative ability among language learners, they overlook the deeper analytical grasp required of language teachers. For educators, syntactic knowledge, as the structural foundation of linguistic competence, is especially important. This expertise encompasses not only the ability to produce grammatically well-formed sentences but also the capacity to analyze the structural mechanisms through which meaning is encoded. This perspective aligns with research in applied linguistics that highlights the importance of pedagogical grammar in teacher preparation, as well as psychometric approaches stressing the need for valid instruments capable of capturing underlying constructs rather than surface performance alone.

Taken together, these insights underscore the necessity of integrating syntactic competence into evaluation frameworks for teacher readiness and professional linguistic expertise. For instance, a single underlying idea can be structurally transformed and expressed in multiple ways depending on speaker's syntactic control. In Chinese, even when the lexical components remain constant, the meaning of a sentence may be shifted depending on word order [33, 41]. Consider the following sentences.[1]

1. a. *They go to school by bus.*
   b. *They take the bus to school.*
2. a. 他們 去學校 搭公車。
   3PL to-go-to-school to-take-bus
   They go to school to take the bus.

[1] Gloss: CL: classifier, DET: determiner, NEG: negation, PROG: progressing, SG: singular, PL: plural, 1, 2, 3: first, second, third person

b. 他們 搭公車 去學校。
3PL to-take-bus to-go-to-school
They take the bus to school.

In examples 1a and 1b, both sentences convey the same idea of how the subjects travel to school. By contrast, sentence 2a refers to the location where the bus is boarded, whereas 2b specifies the mode of transportation used to reach the school. Although both sentences contain identical components, differences in sequential order yield distinct meanings. Another illustration of syntactic ambiguity can be found in the English sentence '*John saw the men with telescope.*' This statement allows for two interpretations: either John is observing the men through a telescope, or John is observing men who themselves possess a telescope. Such prepositional ambiguity demonstrates how structural variation can alter meaning. While native and near-native speakers typically grasp these interpretations with ease, they frequently struggle to articulate the underlying syntactic reasons beyond intuitive understanding. For language teachers, however, the responsibility extends further. They must not only comprehend these ambiguities but also be able to explain the structural mechanisms that generate them. This underscores the argument that syntactic knowledge forms a central component of linguistic competence, providing the analytical foundation necessary for language professionals to move beyond intuitive fluency toward explicit, teachable explanations of how meaning is constructed.

Since these structural rules vary drastically across languages, Chinese is classified as a topic-prominent language [28], standing in contrast to subject-prominent languages like English. Accordingly, sentence structure in Chinese is often organized around a topic-comment relationship rather than a subject-predicate one, reflecting both the absence of obligatory subject-verb agreement and the flexibility of topicalization. As Huang [10] observed, "no agreement between subject and verb is required, the relationship between subject and predicate is often in the topic-comment mode rather than actor-action." For instance, English relies heavily on subject-predicate relations, utilizing strict agreement marking and syntactic constraints foreground the subject's grammatical role. By contrast, Chinese, alongside languages such as Japanese and Korean, exhibit topic-prominent tendencies, where discourse organization privileges topical elements over grammatical subjects.

With reference to this distinction [12], scholars emphasize that topic and subject must be considered as two separate conceptual notions with distinct grammatical properties [28]. Formally defined, a topic in a Chinese sentence is typically placed in the sentence-initial position to declare what the sentence is about. Whereas, a subject is characterized as the entities holding a direct relationship of 'doing' or 'being' with the main verb [27].

3. 那 棵 樹 葉子 很 大
DET CL tree leaf very big
That tree, (its) leaves are very big. [28]

4. 那 棵 樹 葉子 大，我 不 喜歡_____
DET CL tree leaf big 1SG NEG. like
That tree, the leaves are big, (so) I don't like it. [28]

According to discussion, sentence 3 should be regarded as a topic-comment sentence instead of analyzing 那棵樹 '*that tree*' as a main subject and 葉子 '*leaf*' as minor subject [12], as 葉子很大 '*leaves are very big*' comments directly on the tree. This structure establishes a part-whole relationship where the tree, serving as the topic, represents the whole, and the leaves, acting as the subject, represents the part.

Sentence 4 further demonstrates the differences between a topic and a subject. The object of the verb 喜歡 '*like*' refers to the topic 那棵樹 '*that tree*' from the preceding sentence, rather than to the subject 葉子 '*leaves*' [28]. Thus, while a subject shows what a noun phrase is 'doing' or 'being' within a specific clause, a topic usually signals a broader framework, showing what the sentence is about and shaping how information is interpreted across discourse [29, 30,]. In this way, the topic-comment structure serves as an important tool for organizing text.

Since Chinese language relies heavily on topic-prominent structures [9, 10], linear word order becomes a primary principle. Research highlights this fact, asserting that word order in Chinese "is the single most important syntactic device for sentence interpretation" and that "[T]the acquisition of Mandarin syntax cannot begin without knowledge of word order" [10]. This structural reliance on linear sequence governs how heads, complements, and modifiers interact within a sentence. These sentence-level syntactic patterns parallel the internal formation of compound words, taking the forms such as subject-predicate, verb-object, and verb-complement. For instance, the subject-predicate sentence structure seen in sentence 5, where '*he*' acts as the subject performing the action '*to-sleep*', directly mirrors the compound 地震 '*earthquake*' (literally *earth-shake*). Similarly, verb-object structures like 洗衣服 '*washing clothes*' in sentence 6 stand in close contrast to verb-complement structures like 洗乾淨 '*cleanly washed*' in sentence 7.

5. 他 睡覺 了。
3SG to-sleep PTL
He slept.
6. 媽媽 在洗衣服。
mother PROG-to.wash-clothes
Mother is doing laundry.
7. 衣服 洗 乾淨 了。
clothes to-wash clean PTL
The clothes is washed.
8. 他 昨天 來， 今天 還 要 來。 [9]
3SG yesterday to-come today still to-want to-come
He came yesterday, and even today.

9. 颱風 來臨 前 他 買進. 了 許多 的 民生用品。 [9]
typhoon to-come before 3SG to-buy-in PTL many POSS daily necessity
Before typhoon, he bought a lot of essentials.

Following this, sentence 8 exemplifies coordination structures [9], where both parallel constituents are placed side-by-side equally, such as 買賣 '*to trade*', which literally translates to 'buy-sell'. Finally, the modifier-head structure appears in sentence 9, where the phrase '*before typhoon*' acts as a modifier setting a time frame for the subsequent clause '*he bought a lot of essentials*'. It is further elaborated that the Chinese noun phrase 民生用品 '*daily necessity*' follows this same pattern when modified by 許多 '*a lot*', just as the compound word 鐵路 '*railway*' is formed by '*iron*' modifying '*road*', where the first element establishes the category for the second [9].

This structural consistency creates a pedagogical challenge for teachers. Understanding these syntactic structural nuances is essential; while native or near-native speakers may use the language naturally, they are often unable to articulate the underlying rules that govern it. Although they can readily recognize ambiguous sentences and identify subtle variations in meaning, they might struggle to explain why these differences occur. Therefore, for individuals preparing to teach or work in language professions, explicit linguistic competence is indispensable. It provides the analytical foundation necessary to explain syntactic mechanisms rather than relying solely on intuitive fluency.

3 Linguistic Competence and Language Teaching

Linguistic competence is widely recognized as a fundamental construct influencing an individual's ability to express ideas clearly and accurately [33]. It refers to a speaker's internalized system of syntactic, morphological, phonological, and semantic knowledge, which enables the production and interpretation of an infinite number of sentences [13, 14]. Because of this, language teachers play a vital role in the pedagogical profession, serving as the primary models for developing learners' communicative skills. Consequently, a high level of linguistic competence is considered a baseline requirement for language teachers, as inadequate competence can negatively impact student learning outcomes [36, 44]. Research examining the multifaceted nature of language teaching [8, 24, 36, 46] identifies disciplinary knowledge and pedagogical content knowledge as core elements defining a teacher's professional identity.

Professional competence in teaching is a multidimensional construct requiring continuous development. In foreign language education, teachers must demonstrate a distinct domain of proficiency that goes beyond basic communication to encompass academic and specific-purpose language proficiency. Strong linguistic competence enables teachers to convey theoretical concept more precisely, employ appropriate terminology and adjust their explanations to match students' levels [1, 3, 6, 20].

Conversely, limited competence weakens the delivery of content, reducing instructional clarity and credibility even when subject knowledge is present. This misalignment between competence and performance may lead student to question the accuracy of the instruction, causing them difficulty in fully comprehending the material. Linguistic competence contributes to effective teaching by enabling instructors to accommodate learners' needs through demonstrating appropriate teaching methods and providing accurate examples [7]. Thus, explicit linguistic knowledge is particularly vital since the target language itself constitutes the content being taught. While linguistic competence ensures the accuracy of this content, pedagogical content knowledge defines how the content should be delivered [7, 44], shaping instructional clarity and responsive teaching.

Empirical evidence also supports the operationalization of linguistic across modalities. For instance, research investigating instructor's linguistic competence on YouTube [6], demonstrate how core linguistic elements including phonology, morphology, syntax, semantics, and pragmatics, are applied practically to diagnose errors, correct learners' errors, and provide textual explanations. Similarly, in bilingual education, studies emphasized the development of teaching-specific competences [3, 4, 46], such as pedagogical Spanish, as fundamental requirement for academic contexts. Furthermore, a systematic review by Scherzinger & Brahm [38] outlines essential competency frameworks in bilingual education to provide a comprehensive in order to provide a comprehensive model reflecting professional requirements, much of which is grounded in solid theoretical foundations [25]. These studies point out that language instructors must obtain advanced linguistic capabilities to explain content appropriately and draw focus to specific language features [38, 44, 46].

Overall, while linguistic competence does not determine content knowledge of a teacher, it strongly influences how effectively that knowledge is communicated. From a psychometric perspective, previous discussion shows that linguistic competence functions as a latent trait that underpins observable instructional practices. High competence establishes the fundamental proficiency necessary for effective teaching. This helps shaping professional capacity, confidence, and instructional practices in class. Since learners rely on their language instructors as a model of accurate and appropriate language learning sources, linguistic competence is essential in supporting professional identity for potential language teachers.

4 Methodology

A 40-item instrument designed to measure syntactic knowledge in Chinese among prospective Chinese language teachers. The tool assesses six core syntactic structures, which are the Subject-Predicate, Modifier-Head, Verb-Complement, Verb-Object, Coordination, and Resumptive structures. Item design was directly mapped from Chang's and Li and Thompson's framework of Chinese syntactic constituents [9, 28]. Each structure comprises seven items, except for the resumptive structure, which

features five items. Content validity was evaluated and confirmed by a panel of three senior experts in Chinese linguistics.

To evaluate the psychometric properties of the instrument, data were collected from a sample of 112 undergraduate students majoring in Chinese at a university in Taiwan. As prospective Chinese language instructors, many of whom were scheduled to begin classroom teaching the end of the semester. These participants represent a highly relevant population for instrument validation. The instrument was administered digitally during an in-class sessions. Because the digital platform required a response for every item prior submission, the administration yielded zero missing. data.

In order to enhance participant motivation and engagement, items were arranged in an ascending order of difficulty, with easier questions appearing at the beginning and progressively more difficult ones toward the end. Items were structure to ensure that the same correct response option never appeared consecutively more than three times [26]. This design feature was explicitly implemented to minimize guessing behavior associated with the Test Taker's Fallacy, a manifestation of the Gambler's Fallacy [2, 5, 26]. In testing environments, the continuous repetition of identical options often prompts test takers to doubt their knowledge and reconsider their responses, increasing the likelihood that they will arbitrarily shift to a different option [26]. Each item was scored dichotomously with 1 representing a correct response, and 0 representing incorrect response.

The performance data were analyzed using the Rasch model [35], where the probability of test taker *i* correctly answering item *j* is expressed as the following:

$$P(\theta_i) = \frac{exp(\theta_i - b_j)}{1+[exp(\theta_i - b_j)]}. \quad (1)$$

Where θ_i represents the latent trait of test-taker *i* (i.e., syntactic knowledge in this study), and b_j denotes the difficulty parameter of item *j*. Based on Equation (1), the probability of a correct response is highest when item difficulty aligns closely with the test taker's ability level; conversely, when item difficulty exceeds ability, the probability of success decreases.

All Rasch analyses were conducted by the open-source software R [34] with TAM package [37]. Evaluation criteria included fit indices such as unweighted mean square error (Outfit MNSQ) and weighted mean square error (Infit MNSQ). Commonly, MNSQ values between 0.7 and 1.3 indicate acceptable model-data fit [31, 45]. Reliability was assessed using Person Separation Reliability (PSR), which, while conceptually distinct from Cronbach's α, serves a comparable function. A PSR value ≥ .70 was considered acceptable [19]. Finally, a Wright Map was employed to provide a comprehensive visual summary of the measurement system, illustrating the distribution of item difficulties relative to participant abilities.

5 Results and Analysis

Table 1 presents the results of the fit indices, indicating that all items' Infit MNSQ values and the majority of the Outfit MNSQ values fell within the acceptable range of 0.7 to 1.3. A small number of items fell outside these boundaries; specifically, six items exhibited Outfit MNSQ values below 0.7, while three items exceeded 1.3. Generally, elevated MNSQ values greater than 1.3 suggest unpredictability in responses, which may reflect unmodeled noise or model underfit. Conversely, low MNSQ values less than 0.7 indicate overly predictable responses, potentially due to item redundancy or model overfit. Because extreme fit statistics can distort the measurement system, these deviations required careful evaluation. To maximize the robustness and coverage of the developed instrument, we sought to retain as many viable items as possible. Accordingly, the six items with low Outfit MNSQ values were considered acceptable and retained, while the three items with Outfit MNSQ values above 1.3 (Items 6, 13, and 27) were flagged for further qualitative review.

Table 1. Rasch Analysis Result

Item	Difficulty	Standard Error	Outfit MNSQ	Infit MNSQ
1	-2.818	0.363	0.547	0.868
2	-1.396	0.248	0.739	0.851
3	-2.274	0.307	0.972	0.973
4	-1.650	0.262	0.791	0.948
5	-1.789	0.270	1.181	1.003
6	1.483	0.247	1.766	1.230
7	-1.937	0.280	0.851	0.970
8	-1.004	0.232	0.724	0.824
9	-3.103	0.400	0.691	0.880
10	-0.237	0.216	1.111	1.073
11	-0.328	0.217	1.299	1.221
12	-1.279	0.243	1.281	1.221
13	-0.559	0.220	1.341	1.197
14	-1.222	0.240	1.138	1.111
15	-2.098	0.293	1.193	1.042
16	-2.470	0.325	0.622	0.844
17	-0.952	0.231	0.961	0.963
18	-2.470	0.325	0.525	0.785
19	0.386	0.216	0.898	0.940
20	-0.328	0.217	0.933	0.987
21	-1.396	0.248	1.267	1.080
22	-2.016	0.286	0.737	0.925

23	-2.274	0.307	0.456	0.725
24	-1.337	0.246	0.932	0.955
25	-2.274	0.307	0.895	1.026
26	-2.016	0.286	1.040	1.020
27	1.543	0.250	1.462	1.182
28	0.754	0.222	0.876	0.925
29	-2.369	0.316	1.187	0.956
30	-2.016	0.286	0.505	0.761
31	-1.337	0.246	0.957	1.002
32	0.297	0.215	1.187	1.097
33	-1.057	0.234	0.923	0.960
34	0.075	0.214	0.784	0.851
35	-1.004	0.232	0.859	0.970
36	-0.702	0.223	1.165	1.157
37	-2.016	0.286	0.711	0.831
38	-2.274	0.307	0.748	0.892
39	-0.465	0.219	1.147	1.139
40	-0.952	0.231	1.313	1.099

*Note: Outfit MNSQ = unweighted mean square error and Infit MNSQ = weighted mean square error.

As shown in Figure 1, the Item Characteristic Curves (ICCs) for the three flagged items consistently display monotonically increasing patterns with relatively steep slopes, indicating that they satisfy the Rasch model assumption of effective discrimination across varying ability levels. Nevertheless, the performance of test-takers within certain ability ranges did not fully align with model expectations, resulting in the elevated Outfit MNSQ indices. To address this, experts reviewed the content of the three items (as shown below) in relation to the six basic syntactic structures and confirmed that the items were theoretically consistent with these intended structures. Based on this dual consideration of statistical evidence and content validity, all three items were ultimately deemed acceptable for inclusion and retained in the final instrument.

10. 都會 的 繁榮 經濟 吸引 了 很多 人。[Item 6]
metropolis PTL flourishing economy draw PTL many people
The flourishing economy of the city draws many people.

11. 今天 的 比賽 結束 了。[Item 13]
today PTL competition end PTL
The competition today has ended.

12. 聽眾 給予 掌聲 鼓勵。[Item 27]
audience to-give applause to-encourage
The audience gives a round of applause as encouragement.

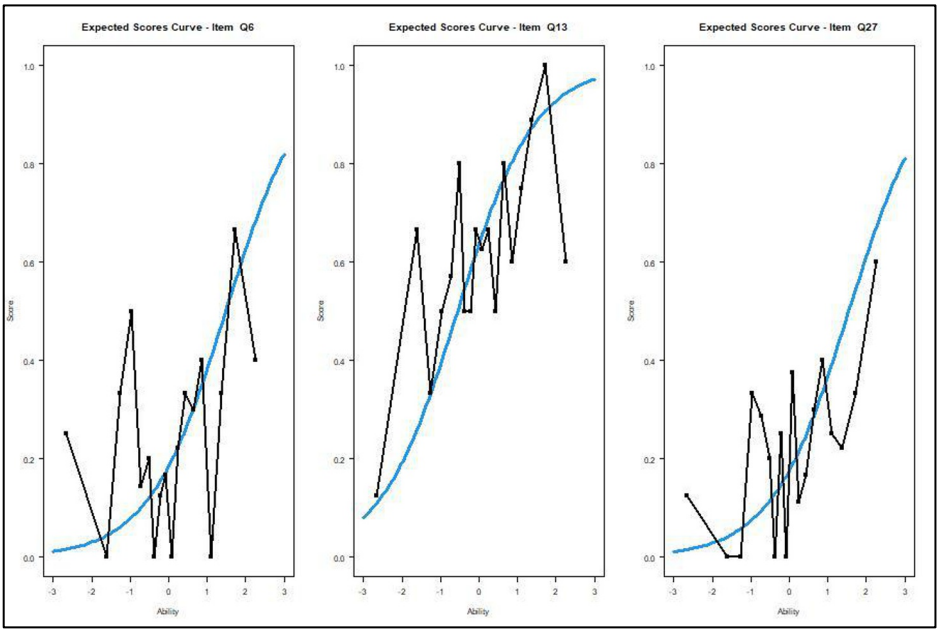


Figure. 1 The Item Characteristic Curves of the Three Mis-fitted Items

Figure 2 illustrates the alignment between item difficulty and test-taker ability using a common logit scale, known as the Wright Map (or Item-Person Map). The Wright Map positions both item difficulty and test-taker ability on the same scale. As seen from Figure 2, the left side displays the distribution of test-taker's ability, ranging from the most capable at the top to the least capable at the bottom. At the right side of the graph illustrates the difficulty distribution of the 40 items, ordered from most difficult to least difficult. The results indicate that the mean ability of the test-takers, centered near 0 logit, closely aligns with the mean item difficulty. Moreover, the ability distribution of over 98% of the test-takers fell within the range of approximately -3 to +3 logits, demonstrating a near-normal distribution. This close alignment suggests that the overall difficulty of the instrument is appropriately matched to the target population, ensuring that the test is neither too easy nor too difficult for prospective instructors.

The PSR of this 40-item instrument was .87, exceeding the commonly accepted threshold of .70. This result indicates that the instrument can be reliably differentiate levels of syntactic knowledge within the sample population.

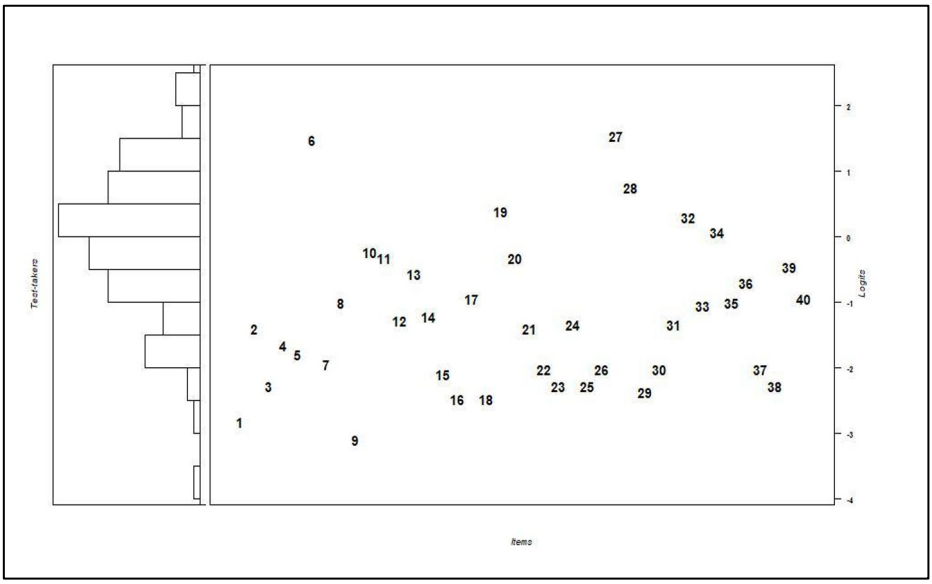


Figure 2. Wright Map of the 40-item instrument

In sum, the examination of fit indices, expert evaluation, ICC analysis, and reliability assessment demonstrated that the developed instrument exhibits a strong fit with the Rasch model. Within the framework of Rasch measurement theory, construct validity is fundamentally established through empirical evidence of data-model fit. In this study, validity evidence is manifested in three key psychometric properties: First, the finding that all 40 items exhibited acceptable Infit MNSQ statistics, with the vast majority falling within the strict 0.7-1.3 threshold for Outfit MNSQ, provides robust empirical verification of unidimensionality. Because the data closely satisfy the prescriptive mathematical requirements of the Rasch model, it demonstrates that all items function cohesively to measure a single, underlying latent trait—Chinese syntactic knowledge—without distortion from unmodeled noise or secondary dimensions. Second, substantive validity is demonstrated via the Wright Map, where the empirical continuum of item difficulties properly mirrors the intended distribution of target population abilities without floor or ceiling gaps. Third, the substantive and qualitative review of the minor misfitting items (Items 6, 13, and 27) by the expert panel confirmed their rigorous alignment with the core theoretical framework. Consequently, these combined psychometric and qualitative results provide compelling evidence that the instrument is both valid and reliable for operationalizing the intended construct across varying ability levels. Accordingly, the instrument can be considered both valid and reliable for measuring the unidimensional construct of syntactic knowledge across a broad range of ability levels.

6 Conclusions and Discussion

This paper presents the initial validation of an assessment framework designed to measure syntactic competence of prospective Chinese language teachers. This instrument specifically targets six phrase structures, including the subject-predicate, verb-objects, verb-compliments, modifier-head, coordination, and resumptive structures. Overall, this study provides strong empirical evidence that the developed 40-item instrument demonstrates robust psychometric properties regarding both construct validity and item-level reliability.

The Rasch analysis offers important baseline insights for assessing students' syntactic knowledge in Chinese. Notably, more than 95% of the items demonstrated difficulty values below -1.0 logits, suggesting the instrument's potential as a diagnostic tool in classroom settings. Such diagnostic use would be particularly valuable for identifying individuals with low to moderate ability levels and for informing instructional adjustments or targeted interventions. However, a closer evaluation of the item difficulty margins reveals a clustering toward the lower-to-middle logit range. To maximize the tools beyond basic diagnostics and prevent potential ceiling effects among highly proficient students, future research needs to pay attention to the margins so that to ensure the test remains valid for high-proficiency students as well as for those with lower ability level. This study provides evidence that the 40-item instrument demonstrates psychometric properties with respect to both construct validity and reliability at the item level.

Despite the promising properties, two primary limitations must be considered when interpretating the findings. First, the sample size of 112 participants represents a key constraint. Although Rasch measurement theory demonstrates that a sample of over 100 participants can yield stable item difficulty calibrations within ±0.5 logits at a 95% confidence level for highly targeted samples [31], a direct psychometric consequence of this modest sample size is the relatively wide standard errors of the item difficulty estimates, which range from 0.21 to 0.40 logits in this study. These higher standard error values suggest that while the relative item hierarchy is established, the exact location parameters require further stabilization. Furthermore, because the participants were drawn from one single institution, selection bias cannot be entirely ruled out, and the generalizability of the item parameters remains restricted.

To address this limitation, future research shall scale up data collection to achieve a larger, more diverse sample size ($N \geq 300$). Expending the sample will not only minimize standard errors of estimation and stabilize item parameter threshold, but also permit advanced psychometric cross-validation, including evaluating Differential Item Functioning (DIF). Testing for DIF across distinct student demographics (e.g., institutional backgrounds, gender, or academic years) will ensure that the assessment framework operates equitably without measurement bias.

Beyond expanding sample size, diversifying the test formats will enable a more versatile evaluation of language competencies. Incorporating open-ended performance tasks alongside traditional multiple-choice and fill-in-the-blank item would facilitate a multidimensional analysis of learner abilities. While closed-ended formats are highly functional for quantitative knowledge retrieval, blending these diverse test formats will provide a more holistic understanding of student capabilities, refine the identification of high-ability individuals, and maintain high measurement validity across all proficiency levels.

From a theoretical standpoint, Chinese differs fundamentally from Indo-European languages in syntax, morphology, and broad grammatical architecture. Because Traditional western models of linguistic competence often rely heavily on inflectional morphology and strict subject-verb agreement paradigms, they may not fully capture the unique structural realities of a topic-prominent language. This potential theoretical mismatch underscores the need for future research to adopt cross-linguistic and cognitive approaches, thereby establishing an analytical framework tailored to Chinese syntactic competence.

Practically, the Wright Map and item difficulty distributions generated through Rasch analysis can serve as valuable diagnostic tools for teachers. The alignment between learner ability and item difficulty allows educators to identify students who may struggle with specific syntactic structures and adjust lesson planning accordingly. For example, items clustered at lower difficulty levels can be used to reinforce foundational knowledge for weaker learners, while more challenging items can be strategically deployed to extend the abilities of advanced students. In this way, psychometric evidence not only validates the instrument but also provides insights for instructional design, supporting differentiated teaching and more effective learning outcomes.

Finally, this study highlights an essential implication for teacher education and professional development. Pre-service teachers frequently rely on the intuitive fluency of their native language and lack of explicit syntactic knowledge required to diagnose their errors or to explain and give instructions clearly. The instruments demonstrated in this study provides instructors with a systematically evaluate tools as well as to strengthen linguistic competence as a measurable construct. This diagnostic approach ensures that teacher candidates not only demonstrating basic communicative proficiency but also acquiring the analytical skills necessary for pedagogical instruction. Embedding psychometrically validated tools into teacher education frameworks can enhance professional identity, build instructional confidence, and prepare language instructors to meet the demands of increasingly diverse and multilingual classrooms.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdulrahman, N. C., Abu-Ayyash, E. A. S.: Linguistic competence, communicative competence and interactional competence. Journal of Advances in Linguistics **10**(1), 1600-1616 (2019)
2. Akyol, P., Key, J., Krishna, K.: Hits or Miss? Test taking behavior in multiple choice exams. Annals of Economics and Statistics **147**, 3-50 (2022)
3. Aquino-Sterling, C. R.: Responding to the call: Developing and assessing pedagogical Spanish competencies in bilingual teacher education. Bilingual Research Journal **39**(1), 50-68 (2016)
4. Aquino-Sterling, C. R., Rodríguez-Valls, F.: Developing Teaching-Specific Spanish Competencies in Bilingual Teacher Education: Toward a Culturally, Linguistically, and Professionally Relevant Approach. Multicultural Perspectives **18**(2), 73-81 (2016)
5. Ayton, P., & Fischer, I.: The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? Memory & Cognition **32**(8), 1369-1378 (2004)
6. Bariq Hammam Rabbani, Sri Sumarni, Ratna Dewanti: Analysing the teacher's application of linguistic competence in English learning and teaching. Teaching English Linguistics Literature Usage (TELL-Us Journal) **9**(2), 413-434 (2023)
7. Barus, I. R. G., Wahyoedi, B.: Investigating Teacher's Linguistic Competence in Teaching English for Academic Purposes. International Journal of English and Applied Linguistics (IJEAL) **2**(3), 348-360 (2022)
8. British Association of Lecturers in English for Academic Purposes: Competency Framework for Teachers of English for Academic Purposes. BALEAP. (2008)
9. Chang, H.-H.: Huayu Jufa Xinlun. ChengChung Book, Taipei (2012)
10. Chang, H.-W.: The Acquisition of Chinese Syntax. Language Processing in Chinese, 277-311 (1992)
11. Chang, L.-P., Tseng, Y.-T.: The effects of an experimental course using data-driven learning approach in Chinese as a second language. Journal of Technology and Chinese Language Teaching **14**(1), 1-25 (2023)
12. Chao, Y.-R.: A grammar of spoken Chinese. University of California Press, Berkeley and Los Angeles (1968)
13. Chomsky, N.: Aspect of The Theory of Syntax. The M.I.T. Press, Cambridge (1965)
14. Chomsky, N.: Language and mind. Harcourt, Brace & World, New York (1968)
15. Chomsky, N.: Rules and representations. Blackwell, Oxford (1980)
16. Chou, C.-T., Chang, L.-P.: A Preliminary Study of Confusable Words in L2 Chinese Textbooks. In M.C. Liu, C.Y. Kit, Q. Su (Eds.), Chinese lexical semantics, pp. 796-809. Springer. (2021)
17. Ellis, R.: The study of second language acquisition. Oxford University Press (1994)
18. Fillmore, C. J., Fillmore, S.: Case Grammar. Universals in Linguistic Theory. Holt, Rinehart, and Winston: New York, 1-88 (1968)
19. George, D., Mallery, P.: IBM SPSS statistics 23 step by step: A simple guide and reference (14th ed.). Routledge (2016)
20. Go, Y.: Error analysis of Chinese word order of Indonesian Students. Humaniora **5**(2), 1140-1145 (2014)
21. Jin, H.-G.: Roles of pushed output in second language acquisition and second language instruction. Shijie Hanyu Jiaoxue **31**(4), 510-541 (2017)
22. Khomenko, O., Buhinska, T., Terletska, L., Hladkoskok, L., Tanana, S.: Formation of linguistic competence of a foreign language teacher (On the example of language) in the system of lifelong learning. Laplage em Revista (International) **7**, 518-527 (2021)

23. Klime, E., Hartig, J., Rauch, D.: The concept of competence in educational contexts. In Hartig, J., Klieme, E., Leutner, D. (Eds.). Assessment of Competencies in Educational Contexts. Hogrefe & Huber Publishers, Gottingen (2008)
24. Krashen, S.: Principles and practice in Second Language Acquisition. Pergamon Press Inc. (1982)
25. Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A.: Professional competence of teachers: Effects on instructional quality and student development. Journal of Educational Psychology **105***(*3), 805–820. (2013)
26. Lee, C. J.: The test taker's fallacy: How students guess answers on multiple-choice tests. Journal of Behavioral Decision Making, 1-12. (2018)
27. Lehmann, C.: Language competence: Theory and empiry. 39th Annual Meeting of the Societas Linguistica Europaea, Bremen. (2009)
28. Li, C.N, Thompson, S.A.: Mandarin Chinese. The Crane Publishing Co., Taipei (2009)
29. Li, Y.-H. A.: Indefinite WH in Mandarin Chinese. Journal of East Asian Linguistics **1**(2), 125-155 (1992)
30. Lin, H.-L.: The interface between syntax and morphology: Taiwanese verbal complexes. Proceedings of the 16th Pacific Asia Conference on Language, Information and Computation (PACLIC 16), 308-319 (2002)
31. Linacre, J. M.: Rasch analysis of rank-ordered data. Journal of Applied Measurement **7**(1), 129-139 (2006)
32. Llurda, E.: On competence, proficiency, and communicative language ability. International Journal of Applied Linguistic **10**(1), 85-96 (2000)
33. Pan, V. J., Paul, W.: The syntax of complex sentences in Mandarin Chinese: a comprehensive overview with analyses. Linguistic Analysis **42**(1-2), (2018)
34. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna Austria. (2012) Retrieved from http://www.Rproject.org/
35. Rasch, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, IL: University of Chicago Press (1960)
36. Richards, C. J.: Competence and Performance in Language Teaching. Cambridge University Press (2011)
37. Robitzsch, A., Kiefer, T., & Wu, M. TAM: Test analysis modules. R package version, **3**, 5-19 (2020)
38. Scherzinger, L., Brahm, T.: A systematic review of bilingual education teachers' competences. Educational Research Review **39**, (2023)
39. Shavelson, J. R.: On the measurement of competency. Empirical Research in Vocational Education and Training **2**(1), 41-63 (2010)
40. Tai, C.-F.: The functions of idioms with parallel construction for foreign students of Chinese language. Journal of Applied Chinese **7**, 289-312 (2010)
41. Tai, J. H-Y.: Quest for the essence of Chinese grammar. The Journal of Chinese Linguistics **51**(3), 582-610 (2023)
42. Taylor, S. D.: The meaning and use of the term 'competence' in linguistics and applied linguistics. Applied linguistics **9**, 148-168 (1988)
43. Tsai, M.-C.: Error categories of cohesion in CSL for Japanese learners. Journal of Chinese Language Teaching, **7**(3), 31-54 (2010)
44. Widdowson, H. G.: Teaching Language as Communication. Oxford University Press, Oxford (1978)

45. Wright, B. D., Linacre, J. M.: Reasonable mean-square fit values. Rasch Measurement Transactions **8**(3), 370-371 (1994)
46. Zúñiga, C. E.: Supporting "Pedagogical" Spanish Language Competencies: Bilingual Teacher Education "en la Frontera". Teacher Education Quarterly **46**(3), 79-98 (2019)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

