

A Study on the C-E Classic Sayings Translation Quality Assessment of LLM DeepSeek-R1 in *Understanding Xi Jinping's Educational Philosophy*

HE Yang^[a]; JIANG Xue^{[b].*}

^[a] Postgraduate, English Department, School of Foreign Languages, North China Electric Power University, Beijing, China.

^[b] Associate Professor, English Department, School of Foreign Languages, North China Electric Power University, Beijing, China.

*Corresponding author.

Received 23 July 2025; accepted 31 August 2025

Published online 26 September 2025

Abstract

This study sampled 30 classic sayings contained in the *Understanding Xi Jinping's Educational Philosophy* as the research objects. It takes the official translation version *Understanding Xi Jinping's Educational Philosophy* which was commissioned to Beijing Foreign Studies University by the National Textbook Committee Office as the reference translation, and the translations generated by the domestically advanced Large Language Model (the following is abbreviated as LLM) DeepSeek-R1 as the objects for translation quality evaluation. The study adopts quantitative evaluation and qualitative analysis as research methods. Through the collaborative analysis of two indicators, namely COMET (Cross-lingual Optimized Metric for Evaluation of Translation) and TER (Translation Edit Rate), it quantitatively evaluates the performance of the evaluation objects in semantic fidelity, and the surface edit distance difference. At the same time, it qualitatively analyzes the cultural-loaded equivalence of the evaluation objects. The study finds that the LLM DeepSeek-R1 can translate well with accurate semantics and suitable contexts in the translation of classic sayings. However, when dealing with expressions with profound cultural connotations, there are still problems such as semantic deviations, insufficient transmission of cultural images, and a measurable surface edit distance from the reference translation. This study provides empirical evidence for enhancing the textual translation quality of the LLM DeepSeek-R1 in the field of classic sayings translation; it also serves as an important reference for future

enhancements of the model in this field and the expansion of its application scenarios.

Key words: LLM DeepSeek-R1; Translation quality assessment; Classic sayings translation; *Understanding Xi Jinping's Educational Philosophy*

He, Y., & Jiang, X. (2025). A Study on the C-E Classic Sayings Translation Quality Assessment of LLM DeepSeek-R1 in *Understanding Xi Jinping's Educational Philosophy*. *Cross-Cultural Communication*, 21(3), 38-47. Available from: <http://www.cscanada.net/index.php/ccc/article/view/13883>
DOI: <http://dx.doi.org/10.3968/13883>

INTRODUCTION

In the era of globalization, Chinese culture and educational concepts are stepping onto the global stage with unprecedented momentum. The *Understanding Xi Jinping's Educational Philosophy* serves as the fundamental guideline for the reform and development of Chinese education in the new era. Its dissemination is of inestimable significance for the world to understand the wisdom of Chinese education. Many classic sayings vividly demonstrate the profound heritage of Chinese education in a concise yet profound way.

Accurate and high-quality English translation is the crucial bridge for the cross-cultural dissemination of these expositions. However, translating classic sayings is not an easy task. The classic sayings carry unique historical and cultural connotations, posing great challenges to translation work. Traditional human translation has certain advantages in ensuring semantic accuracy and the transmission of cultural images. When faced with a large amount of text and the need for rapid dissemination, its efficiency is limited. In recent years, with the rapid development of artificial intelligence technology, LLMs have emerged. They bring new transformative forces to the field of translation.

The domestically advanced LLM DeepSeek-R1 has made a name for itself in the field of natural language processing with its powerful language understanding and generation capabilities. Applying it to the English translation of classic sayings in *Understanding Xi Jinping's Educational Philosophy* is expected to improve translation efficiency and provide a more convenient reading experience for readers worldwide. However, currently, research on the translation quality of LLM DeepSeek-R1 in this specific field is still insufficient. When dealing with classic sayings rich in cultural connotations, questions such as whether it can accurately convey the original semantics, achieve cultural-loaded equivalence, and what differences exist compared with the official reference translation urgently need in-depth exploration.

This study aims to comprehensively evaluate the translation quality of the LLM DeepSeek-R1 in the English translation of classic sayings in *Understanding Xi Jinping's Educational Philosophy* through scientific methods. It is expected to provide empirical evidence for improving the translation quality of LLMs in this field and promote their optimization and application expansion in the translation of classic sayings.

1. OVERVIEW OF LLM DEEPSEEK-R1

In the field of LLMs, prior studies have made substantial progress. Early research focused on the fundamental architecture and training mechanisms of LLMs, with landmark models such as GPT series and BERT laying the theoretical foundation. Scholars explored how these models leverage massive corpora for pre-training to achieve capabilities in language understanding, generation, and translation. For instance, investigations into the GPT series demonstrated their remarkable performance in text generation tasks, while studies on BERT highlighted their strengths in natural language understanding (Devlin, Chang, Lee, & Toutanova, 2019).

As research advanced, attention shifted to the application of LLMs in specific domains. Machine translation emerged as a key area, where numerous studies evaluated the translation quality of LLMs across different language pairs. Regarding domestic LLMs, previous studies have focused on their development and localization adaptability. Researchers have explored how domestic models address language characteristics and cultural nuances specific to Chinese, aiming to improve performance in Chinese-English translation and other tasks (Hu & Li, 2023). DeepSeek-R1, as a representative of domestically advanced LLMs, has drawn attention in recent studies. Preliminary research on DeepSeek-R1 has examined its basic capabilities in text generation and simple translation tasks, showing promising results in terms of fluency and basic semantic accuracy. However,

in-depth studies on its performance in translating classic sayings with profound cultural connotations, especially those related to important political and educational discourses, are still relatively scarce, leaving a research gap that this study intends to fill.

2. OVERVIEW OF TRANSLATION OF CLASSIC SAYINGS IN UNDERSTANDING XI JINPING'S EDUCATIONAL PHILOSOPHY

Research on the translation of classic sayings has a long history. Scholars have explored various translation strategies and principles, emphasizing the importance of balancing semantic accuracy and cultural transmission. Early studies focused on literary classic sayings, discussing how to convey the cultural connotations and aesthetic values embedded in them (Zhang & Zhao, 2025). Translation theories such as domestication and foreignization have been widely applied in analyzing the translation of classic sayings. They provide theoretical guidance for practice (Mo, 2017).

In recent years, with the increasing emphasis on cultural communication, research on the translation of classic sayings in political and ideological discourses has gained momentum. Scholars have examined the translation of key concepts and expressions in important political documents. They highlight the significance of accurate translation in conveying national policies and ideological trends (Feng, 2023). However, specific studies on the translation of classic sayings in *Understanding Xi Jinping's Educational Philosophy* are relatively limited.

Existing research related to *Understanding Xi Jinping's Educational Philosophy* mainly focuses on the interpretation and dissemination of its educational ideas, with little attention paid to the translation of the classic sayings contained within. Some studies have touched upon the translation of educational terms and concepts in the context of China's educational reform (Fu, 2023). But a systematic analysis of the translation quality of classic sayings in this specific work is lacking. This study aims to contribute to this area by evaluating the translation of 30 classic sayings from the book using DeepSeek-R1, building on previous research and addressing the current research deficiency.

3. RESEARCH DESIGN

3.1 Selection of Research Objects

The study takes the official translation *Understanding Xi Jinping's Educational Philosophy* which was commissioned to Beijing Foreign Studies University by the National Textbook Committee Office as the reference

translation. It is published by Foreign Language Teaching and Research Press and Higher Education Press. This version is authoritative and accurate, providing a reliable benchmark for evaluation.

In model selection, this study selects the translations of the domestically advanced LLM DeepSeek-R1 as the core object for analyzing the quality of English translation of classic sayings. Developed by DeepSeek company, a subsidiary of Magic Square Quant, DeepSeek-R1 was globally released on January 20, 2025. It is the first domestic model fully bench marking the GPT-4 architecture with a solid technical foundation. Evaluating the quality of its translations can provide empirical evidence for the performance of LLMs in the field of classic sayings translation, locate their strengths and weaknesses, help optimize domestic models, and support the translation of classic Chinese culture into foreign languages.

The study focuses on evaluating the LLM DeepSeek-R1's translation quality of classic sayings in *Understanding Xi Jinping's Educational Philosophy*. 30 classic sayings from the book are selected as research corpora. These classic sayings contain rich cultural connotations, political wisdom, and educational concepts. These make them ideal samples for evaluating the translation quality of LLMs.

3.2 Construction of Evaluation Framework

This section builds an evaluation framework for classic sayings translation. It will first explain the selection of two quantitative indicators, which include COMET and TER, for measuring translation quality. Then, it focuses on qualitative analysis of cultural-loaded equivalence. Finally, it details the assessment process.

3.2.1 Selection and Explanation of Quantitative Evaluation Indicators

In evaluating translation from Chinese to English, the selection of quantitative evaluation indicators is crucial as they can objectively measure translation quality from different dimensions. This study selects two quantitative evaluation indicators—COMET (Cross-lingual Optimized Metric for Evaluation of Translation) and TER (Translation Edit Rate)—to conduct quantitative analysis on the translation quality of the LLM DeepSeek-R1. The following will analyze these two indicators, expounding their specific roles and characteristics in translation evaluation in this study.

First of all, COMET, whose full name is Cross-lingual Optimized Metric for Evaluation of Translation, is based on a context embedding model and is mainly used to measure semantic fidelity (Maike & Mrinmaya, 2025). Through in-depth analysis of the context information of the source text and the translation, it can evaluate whether the translation is semantically faithful to the original Chinese text. In classic sayings translation, this feature enables it to effectively judge

whether the translation accurately conveys the core semantics of the classic sayings. Its advantage is that it can fully consider the contextual information of the text, thus enabling a more comprehensive evaluation of the accuracy of semantics.

Then it comes to TER, whose full name is Translation Edit Rate, mainly used to measure the surface edit difference between the translation generated by LLM and the reference translation (Snover, & Makhoul, 2006). It is calculated by counting the minimum number of editing operations, which include insertion, deletion, replacement, etc., required to convert one text into another. In this study, by comparing the surface edit distance between the translation generated by DeepSeek-R1 and the reference translation, the degree of difference between them can be intuitively seen. Its advantage is that it can intuitively reflect the surface differences of translations.

In summary, the two quantitative evaluation indicators COMET and TER provide support for evaluating the quality of translation from different angles. Each of them focuses on different evaluation dimensions with unique emphases and application scopes. They can reflect semantic fidelity and surface edit differences of translations to a certain extent. Therefore, in the actual evaluation process, combining these indicators can more comprehensively grasp the quality of classic sayings translation. It provides multi-dimensional references for optimizing translation results.

3.2.2 Qualitative Analysis of Cultural-Loaded Equivalence of Classic Sayings

In the translation of classic sayings in *Understanding Xi Jinping's Educational Philosophy*, qualitative analysis of cultural-loaded equivalence is crucial. Cultural-loaded equivalence requires the translation to mirror not only the literal meaning but also the profound cultural connotations, values, and cultural images embedded in the original Chinese text (Mo, 2017).

Classic sayings are the crystallization of Chinese traditional culture and values. They are full with cultural-specific terms, each carrying unique cultural and historical significance. When translating these sayings, a simple word-for-word translation approach is far from sufficient. Instead, translators need to delve deep into the cultural background and values behind these terms. Meanwhile, educational concepts are also in consideration in this study because of the specific corpus.

In the qualitative analysis, two main aspects are emphasized. The first is the translation of cultural-specific terms. It demands that translators not only seek semantic equivalents but also strive to convey the underlying cultural and educational concepts. These terms are deeply rooted in Chinese culture and educational philosophy. Their accurate translation is essential for the conveyance of the original meaning. The following is the transmission of cultural images. Many classic sayings

employ vivid cultural images to illustrate profound ideas. The translations should ensure that these cultural images are maintained and not distorted. Thus they enable foreign readers to understand and appreciate the original Chinese text's cultural charm.

3.2.3 Assessment Process

Before conducting the evaluation, the translation process must be completed first. Firstly, 30 classic sayings are identified as the objects of translation. These sayings cover different fields and eras, and possess certain representativeness and typicality. Secondly, these 30 classic sayings are input into the DeepSeek-R1 model, with the unified input command "Translate the above content into English" to ensure the consistency of the translation process. After the model generates the translation results, preliminary sorting is carried out to remove obvious format errors and other issues. Thus, translation materials available for evaluation are formed.

After completing the translation process, the evaluation process begins. Firstly, an analysis is conducted on the COMET indicator. Taking the translation results of the 30 classic sayings as the research object, the COMET value of each translated text is calculated, and then the cases are ranked in descending order of the COMET values. This ranking method helps to clearly observe the degree of semantic fidelity between DeepSeek-R1's translations and the reference translations.

Next, an analysis is performed on the TER indicator. The TER value of each translation result is calculated, and the cases are ranked in ascending order of the TER values. This ranking method can intuitively show the differences

between Deepseek-R1's translations and the reference translations at the lexical and structural levels.

Finally, the average values of the translations of the 30 classic sayings in the two indicators of COMET and TER are calculated. By synthesizing the average values of these two indicators and the situations reflected by the ranking of each indicator, a summary is made on the overall performance of DeepSeek-R1 in the translation of classic sayings.

4. EXPERIMENTAL ANALYSIS AND FINDINGS

4.1 Analysis Based on COMET

This section analyzes DeepSeek-R1's classic sayings translations via COMET. It first presents the overall COMET value range to reflect semantic fidelity variation. Then it studies the translation cases of top 5 high-COMET values and bottom 5 low-COMET values, exploring their cultural-loaded equivalence performance. It also links model performance to cultural connotation clarity of the sayings.

4.1.1 Overall Analysis of COMET values

The COMET values for the translations of 30 classic sayings by DeepSeek-R1 range from 0.3598 to 0.8133, indicating a certain degree of variation in the model's semantic fidelity performance. The average COMET value is 0.6300, suggesting the model performs fairly well in accurately conveying the core semantics of classic sayings compared to the reference translations.

Table 1
30 Cases of COMET Value

No.	Chinese classic saying	Official reference translation	LLM DeepSeek-R1's translation	Comet value
1	善之本在教，教之本在师。	The root of goodness lies in education, and the root of education lies in teachers. (p.195)	The foundation of goodness lies in education; The foundation of education lies in the teacher.	0.8133
2	师者，所以传道授业解惑也。	A teacher communicates beliefs, imparts knowledge, and clears confusion. (p.196)	A teacher is one who transmits the way, imparts learning, and resolves doubts.	0.7931
...	...	...	...	...
29	凿井者，起于三寸之坎，以就万仞之深。	He who digs a well starts with a three-inch pit and eventually makes it into a well that is ten thousand feet deep. (p.116)	To govern the world, one must first uphold public-mindedness (fairness); When public-mindedness prevails, the world will be at peace.	0.3906
30	志不求易，事不避难。	Do not set easy goals and never dodge difficulties. (p.60)	The people's livelihood lies in diligence; with diligence, there is no lack.	0.3598
Average	-	-	-	0.6300

4.1.2 Case Analysis of High COMET Values Translations

The following analysis focuses on the five translations with the highest COMET values to explore their performance in terms of cultural-loaded equivalence.

The first case under examination features the original Chinese text “善之本在教，教之本在师。”

. The DeepSeek-R1's translation is “The foundation of goodness lies in education; The foundation of education lies in the teacher.”. It boasts a high COMET value of 0.8133 when compared to the reference translation “The root of goodness lies in education, and the root of education lies in teachers. (p.195)”. From the perspective of cultural-loaded equivalence, this translation excels

both in rendering culture-specific terms and conveying cultural images. “善 (goodness)” carries the moral pursuit advocated by Confucianism. “教 (education)” reflects the emphasis on education. “师 (teacher)” embodies the value of respecting teachers and prioritizing education. They are core concepts in traditional Chinese educational culture. The DeepSeek-R1’s translation accurately pairs these terms with “goodness”, “education”, and “teacher” respectively. They preserve the semantic core of the original words. Meanwhile, they convey the “fundamental and derivative” logical relationship among the three. In doing so, it fully reproduces the cultural connotation of the original Chinese text, which emphasizes the foundational role of education and the importance of teachers.

Next case is with the original Chinese text being “师者，所以传道授业解惑也。” When measured against the reference translation “A teacher communicates beliefs, imparts knowledge, and clears confusion. (p.196)”, the DeepSeek-R1’s translation “A teacher is one who transmits the way, imparts learning, and resolves doubts” achieves a COMET value of 0.7931. This translation accurately captures the core responsibilities of “师 (teacher)”. “道 (the way)” in “传道 (transmits the way)” refers not only to general principles but also encompasses Confucian ethics and life norms. The model effectively conveys the cultural meaning through “transmits the way”. Meanwhile, “授业 (imparts learning)” corresponds to “imparts learning”. At the same time, “解惑 (resolves doubts)” aligns with “resolves doubts”. These three components fully cover the cultural functions of teachers. They include knowledge transmission, ideological guidance, and doubt resolution. So they completely retain the educational and cultural connotations.

The third case involves the original Chinese text “欲筑室者，先治其基。” The DeepSeek-R1’s translation “He

who desires to build a house must first lay its foundation” values a COMET value of 0.7922 when compared to the reference translation “To build a house, first lay the foundations. (p.38)”. The original Chinese text employs the everyday cultural images of “筑室 (building a house)” and “治基 (laying the foundation)”. They are to metaphorize the truth that one must establish a solid base before undertaking any task. Among these terms, “基 (foundation)” denotes not only the physical foundation of a building but also symbolizes the foundation of a career. It carries the pragmatic ideology of “everything succeeds with preparation” in traditional Chinese culture. The DeepSeek-R1’s translation fully preserves the corresponding relationship between the images of “building a house” and “laying the foundation”. It accurately conveys the metaphorical meaning behind these images through “build a house” and “lay its foundation”. This allows readers to grasp the cultural philosophy inherent in the text, achieving equivalent transmission of cultural images.

Then the case’s original Chinese text is “德者，本也。” The DeepSeek-R1’s translation “Virtue is the root” is identical to the reference translation. It earns a COMET value of 0.7857. As a core term in traditional Chinese culture, “德 (virtue)” encompasses multiple connotations, such as morality, character, and integrity. It also serves as the cornerstone of personal cultivation and social order in Confucian thought. “本 (root)” embodies the meaning of “fundamental”, emphasizing the foundational status of “德 (virtue)”. The model uses “Virtue” to correspond to “德” and “root” to correspond to “本”, which are pairs. Thus, it accurately conveys the literal meaning of the original terms. Meanwhile, it profoundly reveals the cultural value of “taking virtue as the root”. So it achieves perfect equivalence for core cultural concepts.

Table 2
Five Cases of Highest COMET Value

No.	Chinese classic saying	Official reference translation	LLM DeepSeek-R1’s translation	Comet value
1	善之本在教，教之本在师。	The root of goodness lies in education, and the root of education lies in teachers. (p.195)	The foundation of goodness lies in education; The foundation of education lies in the teacher.	0.8133
2	师者，所以传道授业解惑也。	A teacher communicates beliefs, imparts knowledge, and clears confusion. (p.196)	A teacher is one who transmits the way, imparts learning, and resolves doubts.	0.7931
3	欲筑室者，先治其基。	To build a house, first lay the foundations. (p.38)	He who desires to build a house must first lay its foundation.	0.7922
4	德者，本也。	Virtue is the root. (p.43)	Virtue is the root.	0.7857
5	师者，人之模范也。	Teachers are models of good conduct. (p.198)	Teachers are models for people.	0.7849

Last to detail is the fifth case, which the original Chinese text reads “师者，人之模范也。” When compared to the reference translation “Teachers are models of good conduct. (p.198)”, the DeepSeek-R1’s

translation “Teachers are models for people” achieves a COMET value of 0.7849. The term “模范 (model)” in the original Chinese text carries the expectation in traditional Chinese culture that teachers should serve as moral

examples and behavioral role models. It embodies the educational concept that “a teacher should be both erudite and upright”. The DeepSeek-R1’s translation accurately conveys the cultural connotation of “模范” through “models for people”. It highlights the exemplary role of teachers in society. This enables readers to understand the high emphasis placed on teachers’ conduct in traditional Chinese education.

These high COMET-value translations demonstrate the model’s strong capacity for semantic fidelity in handling classic sayings with clear semantic structures and relatively straightforward cultural connotations. The model successfully grasps the meaning of culture-specific terms. Then it conveys the embedded cultural images and values effectively. This advantage stems from the concise expression and explicit cultural orientation of the original Chinese text. It allows the model to more accurately match corresponding cultural elements in the target language and accomplish equivalent transmission. The detailed statistics are shown in the Table 2.

4.1.3 Case Analysis of Low COMET values Translations

The following analysis focuses on the five translations with the lowest COMET values to explore their deficiencies in terms of cultural-loaded equivalence.

Opening the analysis is the case of “志不求易，事不避难。” The DeepSeek-R1’s translation is “The people’s livelihood lies in diligence; with diligence, there is no lack.”. It attains a COMET value of merely 0.3598 against the reference “Do not set easy goals and never dodge difficulties. (p.60)”. Culturally, this output exhibits substantial divergence in interpreting culture-specific terms and transmitting connotations. The original term “志 (aspiration)” embodies a core traditional concept urging the pursuit of lofty ideals, while “事不避难 (never dodge difficulties)” reflects values of resilience and responsibility. The translation incorrectly links these to “people’s livelihood” and “diligence”, completely detaching from the original cultural context. This results in a failure to convey the emphasis on proactivity and courage. Thus it leads to inadequate cultural-loaded equivalence.

Proceeding to the next example, “凿井者，起于三寸之坎，以就万仞之深。”, DeepSeek-R1 generates “To govern the world, one must first uphold public-mindedness (fairness); When public-mindedness prevails, the world will be at peace.”, scoring 0.3906 in COMET compared to the reference “He who digs a well starts with a three-inch pit and eventually makes it into a well that is ten thousand feet deep. (p.116)”. The original centers on the cultural image of “digging a well”, symbolizing perseverance and incremental effort. The DeepSeek-R1’s version misinterprets this image, associating it instead with unrelated notions of governance and public-

mindedness. This distortion undermines both the cultural image and its underlying philosophy, rendering cultural-loaded equivalence virtually absent.

Another instance is “国有贤良之士众，则国家之治厚；贤良之士寡，则国家之治薄。”. Here, DeepSeek-R1 outputs “Learning is like a crossbow, talent like an arrowhead. Only when guided by insight (or knowledge) can the arrow hit the bull’s-eye.”, with a COMET value of 0.4340 against the reference “When the virtuous are numerous in the state, order will be stable; when the virtuous are few, order will be unstable. (p.56)”. The original discusses “贤良之士 (virtuous and talented people)”, a key concept in traditional political culture, emphasizing merit-based governance. DeepSeek-R1 introduces irrelevant metaphors of learning and archery, deviating entirely from the intended meaning. This failure to accurately render culture-specific terms or convey the core cultural message results in a significant lack of cultural equivalence.

Further examining “民生在勤，勤则不匮。”, DeepSeek-R1 offers “He who grasps the greater principle can also encompass the lesser details.”, achieving a COMET value of 0.4619 relative to the reference “Poverty is a stranger to industry. (p.60)”. The original links “people’s livelihood (民生)” with “diligence (勤)”, conveying a cultural belief in hard work as foundational to prosperity. Although the translation shows some semantic connection, it misses the precise meanings of these culture-specific terms and fails to express the logical relationship between diligence and livelihood, leading to a notable shortfall in cultural transmission.

Finally is “与人不求备，检身若不及。”. DeepSeek-R1 generates “Towards others, do not demand perfection; examine yourself as if you always fall short.”, scoring 0.4802 on the COMET value compared to “Be strict with oneself and lenient with others. (p.39)”. While partially capturing the literal meaning, the translation lacks precision and fluency in conveying the Confucian values of tolerance and self-reflection. The moral depth behind “tolerance” and “self-examination” is insufficiently emphasized, hindering readers’ understanding of these traditional ethical concepts and indicating limited cultural-loaded equivalence.

These low COMET-value translations reflect DeepSeek-R1’s limitations in semantic fidelity when processing classic sayings that carry profound cultural connotations. The model struggles to accurately capture the meaning of culture-specific terms and effectively convey the underlying cultural images and values. This leads to inadequate performance in achieving cultural-loaded equivalence. This is primarily due to the implicit cultural orientation of such original Chinese texts and the highly unique nature of their cultural terms and images. The detailed statistics are shown in the following Table 3.

Table 3
Five Cases of Lowest COMET Value

No.	Chinese classic saying	Official reference translation	LLM DeepSeek-R1's translation	Comet value
1	志不求易，事不避难。	Do not set easy goals and never dodge difficulties. (p.60)	The people's livelihood lies in diligence; with diligence, there is no lack.	0.3598
2	凿井者，起于三寸之坎，以就万仞之深。	He who digs a well starts with a three-inch pit and eventually makes it into a well that is ten thousand feet deep. (p.116)	To govern the world, one must first uphold public-mindedness (fairness); When public-mindedness prevails, the world will be at peace.	0.3906
3	国有贤良之士众，则国家之治厚；贤良之士寡，则国家之治薄。	When the virtuous are numerous in the state, order will be stable; when the virtuous are few, order will be unstable. (p.56)	Learning is like a crossbow, talent like an arrowhead. Only when guided by insight (or knowledge) can the arrow hit the bull's-eye.	0.4340
4	民生在勤，勤则不匮。	Poverty is a stranger to industry. (p.60)	He who grasps the greater principle can also encompass the lesser details.	0.4619
5	与人不求备，检身若不及。	Be strict with oneself and lenient with others. (p.39)	Towards others, do not demand perfection; examine yourself as if you always fall short.	0.4802

4.1.4 Implications for Improvement of LLM DeepSeek-R1

Based on the analysis of COMET values, it is clear that DeepSeek-R1 needs to improve its ability to understand and translate classic sayings with deep cultural and semantic meanings. This may require integrating more in-depth cultural and historical knowledge into its training data. For example, this requires detailed explanations of the philosophical concepts and political ideas behind these sayings. Moreover, the model could benefit from more advanced algorithms that can better analyze the context and semantic relationships in these complex sayings. So it can be improved in semantic fidelity.

4.2 Analysis Based on TER

This section analyzes DeepSeek-R1's classic sayings translations using TER. It first presents the overall TER

value range to show lexical and structural differences from references. Then it studies the translation cases of top 5 low-TER values and bottom 5 high-TER values, exploring vocabulary and structure consistency. Finally, it offers implications to enhance the model's alignment with reference translations.

4.2.1 Overall Analysis of TER Values

The TER values for DeepSeek-R1's translations of 30 classic sayings range from 0.0000 to 1.7778. This indicates significant differences between the translated texts and the reference translations at the lexical and structural levels. The average TER value is 0.9010, which shows that, on average, there are relatively obvious differences in vocabulary and structure compared with the reference translations, and the overall closeness of the translations needs to be improved.

Table 4
30 Cases of TER Value

No.	Chinese classic saying	Official reference translation	LLM DeepSeek-R1's translation	TER value
1	德者，本也。	Virtue is the root. (p.43)	Virtue is the root.	0.0000
2	人而无德，行之不远。	A person without virtue will not go far. (p.44)	A person without virtue will not go far in conduct.	0.2222
...	...	...	...	...
29	民生在勤，勤则不匮。	Poverty is a stranger to industry. (p.60)	He who grasps the greater principle can also encompass the lesser details.	1.7143
30	与人不求备，检身若不及。	Be strict with oneself and lenient with others. (p.39)	Towards others, do not demand perfection; examine yourself as if you always fall short.	1.7778
Average	-	-	-	0.9011

4.2.2 Case Analysis of Low TER Values Translations

The top five translations with the lowest TER values are as follows.

Beginning with the maxim “德者，本也。”，the DeepSeek-R1's translation “Virtue is the root” achieves perfect alignment with the reference “Virtue is the root (p.43)”, reflected in a TER value of 0.0000. The simplicity

of this saying, which is with its direct subject-predicate structure and clear semantic link between “virtue” and “root”, allows the model to demonstrate exceptional accuracy in both vocabulary selection and structural reproduction.

Moving to the second example, “人而无德，行之不远。”，the model outputs “A person without virtue will

not go far in conduct”, which yields a TER of 0.2222 against the reference “A person without virtue will not go far (p.44)”. While both versions maintain the core logic that virtue deficiency limits progress, the model expands slightly by adding “in conduct” to specify the meaning of “行”. Moreover, the fundamental structure, which expresses “person without virtue” leads to “not far”, remains intact, preserving the original intent.

Turning next to “善之本在教，教之本在师。”, the DeepSeek-R1’s translation is “The foundation of goodness lies in education; The foundation of education lies in the teacher”, scoring 0.3529 in TER compared to the reference “The root of goodness lies in education, and the root of education lies in teachers (p.195)”. Here, the model substitutes “root” with the near-synonym “foundation” and uses the singular “teacher” instead of the plural form. Structurally, it employs two independent clauses joined by a semicolon, whereas the reference combines them with “and”. Nevertheless, both versions faithfully retain the hierarchical relationship of “goodness → education → teacher”, differing only in stylistic choices.

Examining the fourth case, “师者，人之模范也。”, the DeepSeek-R1’s translation “Teachers are models for people” achieves a TER of 0.4286 relative to the reference “Teachers are models of good conduct (p.198)”. The model’s phrasing, “models for people”, broadens

the scope compared to the reference’s focus on “good conduct”. Structurally, both follow the “Teachers are. . .” pattern, though the reference adds specificity by qualifying “models”. Despite this difference, the central idea that teachers serve as exemplars is effectively communicated, with deviations limited to minor lexical and structural variations.

Concluding the analysis is the “苟日新，日日新，又日新。”. The model renders it as “If you can renew yourself one day, do so each day, and again each following day”, attaining a TER of 0.4545 against the reference “If you can one day renovate yourself, do so from day to day, and let there be daily renovation (p.184)”. While the model uses “renew” instead of “renovate” and structures the clauses sequentially with “do so. . . again. . .”, the reference opts for “from day to day” and “let there be. . .”. Despite these subtle divergences in phrasing and syntax, the DeepSeek-R1’s translation accurately conveys the original’s emphasis on continual self-renewal.

These translations with low TER values reflect the model’s capability in handling classic sayings characterized by simple semantics and straightforward structures. The model demonstrates an ability to maintain strong consistency in vocabulary choice and structural organization. The detailed statistics are shown in the following Table 5.

Table 5
Five Cases of Highest TER Value

No.	Chinese classic saying	Official reference translation	LLM DeepSeek-R1’s translation	TER value
1	德者，本也。	Virtue is the root. (p.43)	Virtue is the root.	0.0000
2	人而无德，行之不远。	A person without virtue will not go far. (p.44)	A person without virtue will not go far in conduct.	0.2222
3	善之本在教，教之本在师。	The root of goodness lies in education, and the root of education lies in teachers. (p.195)	The foundation of goodness lies in education; The foundation of education lies in the teacher.	0.3529
4	师者，人之模范也。	Teachers are models of good conduct. (p.198)	Teachers are models for people.	0.4286
5	苟日新，日日新，又日新。	If you can one day renovate yourself, do so from day to day, and let there be daily renovation. (p.184)	If you can renew yourself one day, do so each day, and again each following day.	0.4545

4.2.3 Case Analysis of High TER Values Translations

The bottom five translations with the highest TER values are focused in this part.

First, the model translates the classic saying “与人不求备，检身若不及。” as “Towards others, do not demand perfection; examine yourself as if you always fall short.” When this model-generated translation is contrasted with the reference translation “Be strict with oneself and lenient with others (p.39)”, its TER value is 1.7778. From the vocabulary perspective, the DeepSeek-R1’s translation uses “demand perfection” and “fall short” to replace “strict with oneself” and “lenient with others” in the reference translation. Among these replaced phrases, “strict with oneself” in the reference translation conveys the meaning of self-discipline, while “lenient with others” conveys the meaning of tolerance for others. This

vocabulary replacement in the DeepSeek-R1’s translation results in a shift in the semantic focus of the original classic saying. Structurally, the DeepSeek-R1’s translation splits into two separate clauses that center on “others” and “self” respectively, while the reference translation adopts a concise contrast structure to connect self-strictness and others-lenience.

Next, regarding “民生在勤，勤则不匮。”, the DeepSeek-R1’s translation is “He who grasps the greater principle can also encompass the lesser details”, and when compared with the reference translation “Poverty is a stranger to industry (p.60)”, its TER value reaches 1.7143. Semantically, the DeepSeek-R1’s translation abandons the original logical link between livelihood (民生), diligence (勤) and no scarcity (不匮). Vocabulary such as “greater principle” and “lesser details” in the DeepSeek-R1’s

translation is irrelevant to the core concepts of “livelihood” and “diligence” in the original sentence. Structurally, the DeepSeek-R1’s translation uses a “who...can...” clause that focuses on abstract principles, which differs from the reference translation. This translation adopts direct “poverty-industry” logic to reflect the relationship between livelihood and diligence.

Moving on, the model translates “志不求易，事不避难。” as “The people’s livelihood lies in diligence; with diligence, there is no lack”, and when set against the reference translation “Do not set easy goals and never dodge difficulties (p.60)”, its TER value is 1.5000. In terms of vocabulary, the model misinterprets “志” and “易”: “志” originally means ambition, but it is taken as “livelihood” in the DeepSeek-R1’s translation; “易” originally means easy tasks, but it is regarded as “diligence” instead. Structurally, the DeepSeek-R1’s translation shifts to an unrelated theme of “livelihood-diligence” and gives up the “do not...never...” parallel structure used in the reference translation. This structure in the reference is meant to express the ideas of “goals” and “difficulties”.

Subsequently, for the sentence “水之积也不厚，则其负大舟也无力。”, the DeepSeek-R1’s translation is “If the accumulation of water is not deep, it cannot float a large vessel”, and when compared with the reference translation “Shallow water cannot lift a big ship (p.200)”, its TER value is 1.3333. In vocabulary aspect, the DeepSeek-R1’s translation uses “accumulation...”

not deep” and “float a vessel”, whereas the reference translation uses “shallow water” and “lift a ship” instead. Structurally, the DeepSeek-R1’s translation adopts a conditional “if...cannot...” clause, while the reference translation uses a more concise direct statement form.

Lastly, the model translates “苟利于民，不必法古；苟周于事，不必循俗。” as “If it benefits the people, there’s no need to rigidly follow antiquity; If it accomplishes the task, there’s no need to rigidly adhere to convention”, and when contrasted with the reference translation “Follow not the beaten path to bring benefits to the people; observe not old conventions to get things done (p.146)”, its TER value is 1.2381. In terms of vocabulary, the DeepSeek-R1’s translation uses “rigidly follow” and “accomplishes the task”: “rigidly follow” emphasizes the quality of rigidity, and “accomplishes the task” has a narrower scope than “get things done” in the reference translation. Structurally, the DeepSeek-R1’s translation uses two “if...there’s no need...” clauses, which is different from the forceful “follow not...observe not...” parallel structure in the reference translation.

These translations with high TER values underscore the model’s limitations in dealing with classic sayings that possess certain cultural connotations, complex semantics, or unique expression styles, showing a significant surface edit distance difference from the reference. This indicates its insufficient ability to accurately reproduce the original text at both the lexical and structural levels. The detailed statistics are shown in the following Table 6.

Table 6
Five Cases of Lowest TER Value

No.	Chinese classic saying	Official reference translation	LLM DeepSeek-R1’s translation	TER value
1	与人不求备，检身若不及。	Be strict with oneself and lenient with others. (p.39)	Towards others, do not demand perfection; examine yourself as if you always fall short.	1.7778
2	民生在勤，勤则不匮。	Poverty is a stranger to industry. (p.60)	He who grasps the greater principle can also encompass the lesser details.	1.7143
3	志不求易，事不避难。	Do not set easy goals and never dodge difficulties. (p.60)	The people’s livelihood lies in diligence; with diligence, there is no lack.	1.5000
4	水之积也不厚，则其负大舟也无力。	Shallow water cannot lift a big ship. (p.200)	If the accumulation of water is not deep, it cannot float a large vessel.	1.3333
5	苟利于民，不必法古；苟周于事，不必循俗。	Follow not the beaten path to bring benefits to the people; observe not old conventions to get things done. (p.146)	If it benefits the people, there’s no need to rigidly follow antiquity; If it accomplishes the task, there’s no need to rigidly adhere to convention.	1.2381

4.2.4 Implications for Improvement of LLM DeepSeek-R1

Based on the analysis of TER values, it is clear that DeepSeek-R1 needs to improve its ability to keep the

vocabulary and structure close to the reference translations when translating classic sayings with cultural connotations and complex semantics. This may require adding more examples of classic sayings of different types and cultural

backgrounds to its training data, and conducting in-depth analysis and learning of the vocabulary characteristics and structural rules of these sayings. In addition, the model can introduce more advanced vocabulary and structure matching algorithms to better select appropriate vocabulary and construct proper sentence structures during the translation process, thereby reducing the differences in vocabulary and structure from the reference translations and improving the translation quality.

5. CONCLUSIONS

5.1 Research Summary

This study evaluates the quality of LLM DeepSeek-R1 in translating Chinese classic sayings related to *Understanding Xi Jinping's Educational Philosophy* into English. It uses quantitative indicators including COMET and TER to evaluate the translation quality. The analysis reveals that Deepseek-R1 demonstrates varying performance across different classic sayings. It achieves high accuracy in translating those with simple structures and clear semantic relationships. They are evidenced by high COMET values and low TER values. However, for sayings with profound cultural connotations, complex semantic structures, or unique idiomatic expressions, the model shows significant limitations. It shows problems like semantic deviations and notable surface edit differences from reference translations. These findings collectively reflect the model's strengths and weaknesses in translating classic sayings within the specific context of educational philosophy.

5.2 Core Contributions

This research contributes to the field in several aspects. Firstly, it focuses on the understudied area of classic sayings translation related to *Understanding Xi Jinping's Educational Philosophy*, providing empirical data on the performance of the LLM DeepSeek-R1 in this specific domain, which enriches the research on LLM-based translation quality assessment. Secondly, by integrating multiple quantitative evaluation indicators, it offers a practical reference for similar studies on translation quality evaluation of cultural and philosophical texts. Thirdly, the identified strengths and weaknesses of DeepSeek-R1 in handling such translations provide insights for optimizing LLMs' ability to translate classic sayings with cultural and philosophical significance. Thus, it helps to bridge the gap between theoretical research on LLM translation and practical application in cultural communication.

5.3 Research Limitations and Prospects

This study also has limitations. First of all, the research

object is only 30 classic sayings from *Understanding Xi Jinping's Educational Philosophy*, restricting the generalizability of conclusions. Moreover, the evaluation relies on only two indicators, it may be not so comprehensive. Future research can expand the sample sources and size and use more indicators for evaluation. As for optimizing LLMs, researchers can do targeted training based on existing limitations to improve the translation quality of classic sayings.

REFERENCES

- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Retrieved from <https://doi.org/10.48550/arXiv.1810.04805>.
- Feng, M. (2023). *Translation for China's Global Communication from the Perspective of Political Equivalence: A Case Study of Understanding Xi Jinping's Educational Philosophy*. Retrieved from <https://doi.org/10.26962/d.cnki.gbjwu.2023.000150>.
- Fu, T. J. (2023). *Knowledge Construction and Discourse Shaping in Translation Education: With Reference to the English Translation of The Excerpts from Xi Jinping's Important Discourses on Education*. Retrieved from <https://doi.org/10.26914/c.cnkihy.2023.114163>.
- Hu, K. B. & Li, X. Q. (2023). The development of translation studies in the era of large language models: Challenges and prospects. *Chinese Translators Journal*, 44(06), 64-73+192.
- Maike, Z., & Mrinmaya, S. (2025). *COMET-poly: Machine Translation Metric Grounded in Other Candidates*. Retrieved from <https://doi.org/10.48550/arXiv.2508.18549>.
- Mo, C. (2017). On the English Translation of Characteristic Words in Chinese Classics from the Perspective of Equivalent Translation Theory. *Journal of Suzhou University*, 32(11), 62-65.
- Mo, J. (2017). Discussion on the Application of Foreignization in News Translation-Han Shan Yi, Han Jie Fang Jun Nan as Example-Han Shan Yi, Han Jie Fang Jun Nan as Example. *Journal of Beijing Institute of Graphic Communication*, 25(05), 27-29.
- Snover, M.G., & Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. *Conference of the Association for Machine Translation in the Americas*. Retrieved from <https://api.semanticscholar.org/CorpusID:263887736>.
- Zhang, S. K., & Zhao, C. Y. (2025). A Study on the Appropriateness of Large Language Models in Literary Translation: A Comparative Analysis of the Translation Quality of Multiple Models for Border Town Based on Multi-Indicator Evaluation. *Foreign Languages in China*, 22(04), 85-95.