# How do Chinese students' critical thinking compare with other students?: a structured review of the existing evidence

**Keji Fan[1, *], Beng Huat See[1]**

[1.] *School of Education, Durham University, Durham, DH1 1TA, United Kingdom*

*Correspondence: keji.fan@durham.ac.uk

# How do Chinese students' critical thinking compare with other students?: a structured review of the existing evidence


## Abstract

An increasing number of Chinese students are now studying abroad in western universities, and there is a widespread concern among western academics that Chinese students are not trained to have a critical mind. However, there is little empirical evidence so far as to whether this is actually the case. This paper presents the results of a systematic review of international studies that compare the critical thinking of Chinese students with students of other nationalities. A search of eight social science databases supplemented by other sources found 15 studies that met pre-specified inclusion criteria. Nine of these focused on students' critical thinking skills, but their results were mixed. There is no evidence to support the claim that Chinese students have higher or lower critical thinking skills than other students. The research in this area is weak. Five studies on critical thinking dispositions suggest that Chinese students were less disposed to critical thinking, which is not the same as being weak in critical thinking. Only one study was about critical thinking style, indicating that Chinese students are better at information-seeking than peers in other countries. All studies were small-scale using weak designs. These findings suggest that the critical thinking of Chinese students is under-studied, and therefore, more robust, larger-scale experimental studies are needed.



**Keywords:** critical thinking; Chinese students; systematic review

## 1. Background

As the number of Chinese students studying overseas in western universities has increased in the last decade or more, there is a growing interest in the learning skills and dispositions of Chinese students. One common stereotypical perception of Chinese students is that they are somehow deficient in critical thinking (Song, 2014; Xu, 2021). Lucas (2019), for example, suggested that Chinese learners lack CT skills training such as analysing and evaluating information. Guo and O'Sullivan (2012) observed that Chinese students were not familiar with critical thinking (CT) and misunderstood it as negative thinking. The word 'critical' is often interpreted to mean to 'criticise' in the Chinese language. This has led to a misunderstanding that to be "critical" means to be rude. Chinese students have also been reported to face challenges in clarifying their ideas in international class discussions (Guo & O'Sullivan, 2012). Chinese students are portrayed as passive recipients of knowledge (Lucas, 2019). Their learning is superficial (Watkins & Biggs, 1996), focusing on memorisation rather than interpretation or analysis.

The Chinese culture of conformity and respect and reverence for authority perhaps explain their reluctance to question and to argue. But this is not to say that they are less adept at critical analysis, although it is often interpreted as such by academics in western democracies. The stereotype image of Chinese students also perhaps stems from Atkinson's conceptualisation of CT and an inappropriate measure to assess CT in many previous studies. Under the influence of the Confucian culture, students are educated to value conformity (Watkins & Biggs, 1996), and not to question authority (Paton, 2005). According to Atkinson (1997), CT is essentially embedded in western cultures. As a distinct and unique product in western culture, CT is incompatible with Asian culture. This conceptualisation implies that Chinese students naturally lack CT. Unfamiliarity with western academic traditions offers another explanation for Chinese students' poor CT (Paton, 2005; Turner, 2006). For those who come to study in the UK for the first time, unfamiliarity with western academic traditions such as academic writing style, where higher level of critical analysis and ability to present opposing viewpoints are expected, their safe, uncritical and unsceptical review of literature, for example, may be taken as a demonstration of lack of critical thinking (Turner, 2006).

English language proficiency may also be a barrier to understanding critical thinking tests questions. Most measurements of CT are developed by researchers in the west (e.g., the Watson Glaser Critical Thinking Test, the California Critical Thinking Test and the Cornell Critical Thinking Test) where English is the language of the assessment. A certain degree of English language proficiency may be needed for such assessments (Moosavi, 2021). Such factors are often not considered when assessing CT of people of different nationalities. Overall, because of educational, cultural and linguistical reasons, Chinese students have been portrayed as not critically skilled or disposed as their foreign counterparts.

This stereotypical view is so entrenched that many studies have accepted it and tried to investigate reasons for the lack of CT among Chinese students (e.g., Durkin, 2011; Paton, 2005; Zhang, 2017). Guo and O'Sullivan (2012) and Lucas (2019), for instance, recognised that the ambiguity of CT may cause confusion about what it is precisely. These studies may unwittingly reinforce the stereotype (Moosavi, 2021), which may lead Chinese students to internalise the negative discourse on their deficit ability in CT (Song, 2014; Xu, 2021). Thus, Chinese students are less confident and vocal in voicing their opinions, further reinforcing western academic's perception of them as critically unaware (Li, Chen, & Duanmu, 2010). Consequently, some western academics have taken this stereotype for granted and have tried to design tailored

curricula for Chinese students (Badger, 2019). Few have tried to establish the CT skills of Chinese students before accepting these general stereotypical views. This is important as efforts and money will be wasted in designing interventions to improve the CT skills of Chinese students if there is no evidence that Chinese students lack CT skills. We would be solving a problem that does not even exist in the first place.

Indeed, some scholars have attempted to challenge this image (e.g., Heng, 2018; Li, 2013; Lu & Singh, 2017), but they do so by interpreting CT in the Chinese context (Lu & Singh, 2017) or exploring evidence of CT in Chinese students' learning (Li, 2013). In many cases, however, the judgement of Chinese students' CT is based on subjective impressions, which is notoriously unreliable. We would not measure students' maths ability by asking their teachers' opinions, neither would we test their maths ability using a foreign language, so why would we measure Chinese students' CT skills using these measurements. Sometimes the English language proficiency may also be misused as an aspect to measure CT (Moosavi, 2021). A more reliable evaluation of CT skills would be the use of standardised tests (Gorard, See, & Siddiqui, 2017). Few studies used standardised tests to measure Chinse students' critical thinking. And the few that did simply measure the CT skills level of students without any comparison. With no comparisons with students from other nationalities, it is not possible to conclude whether Chinese students have higher, lower or comparable CT skills than say similar students in western democracies (Gorard, 2013). The assumption that Chinese students lack CT/ are deficient/ poor in CT implies a comparison. What are we comparing Chinese students' CT with, whose CT are we comparing and what does the norm look like? Most research into students' CT does not have a comparator, and yet made bold claims about the low levels of CT skills of Chinese students. This is absurd, and yet widely accepted.

For this reason, our research is a review of credible studies that compare Chinese students' CT with that of other nationalities using validated standardised tests.

## 2. Theoretical understanding of critical thinking

### *2.1 What is critical thinking*

Critical thinking is a contentious term (Byrne, 1994; Fisher, 2011; Nilson, 2021). To illustrate, Ennis (1987) argues that CT assists individuals to justify beliefs or actions by rational and self-regulatory thoughts. Lipman (2003), on the other hand, suggests that less attention should be given to the outcome of thinking and more to the specific process of thinking. The same problem exists in the definition from Barnett and Davies (2015). They emphasize the educational importance of CT and argue that it enables students to achieve their potential. To gain a precise definition of CT, Facione (1990) invited scholars from various disciplines (e.g., philosophy, psychology, education) to define CT. The result of this Delphi project revealed that CT is 'purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological criteriological, and contextual considerations upon which that judgment is based' (p. 3).

Despite conceptual variations in the literature on CT, there are parallels in the essential components of CT in various thinking skills frameworks (see Black, 2012; Dwyer, Hogan, & Stewart, 2014; Facione, 1990; Nilson, 2021). The overlaps among CT skills include a) analysis, which may aid in the investigation, examination, and identification of the propositions within an argument (Dwyer et al., 2014); b) evaluation, which may be used to examine an argument

in terms of validity, reliability, relevance, and the possibility of biases (Dwyer et al., 2014); c) inference, which entails gathering valid, trustworthy, and relevant evidence that leads to a proper conclusion (Dwyer et al., 2014).

Another dimension to define CT is affective dispositions. The CT dispositions are related to internal inclination towards final decisions or actions (Facione, Sánchez, Facione, & Gainen, 1995). Researchers have revealed several important propensity elements of CT (see Bailin, Case, Coombs, & Daniels, 1999; Ennis, 1985; Facione, 1990; Halpern, 1998) such as open-mindedness, truth-seeking and inquisitiveness. Notably, dispositions and skills are different. For example, a licensed driver might not be willing to drive a car. Likewise, individuals being skilled in CT may not possess positive CT characters and vice versa (Facione, Facinoe, & Giancarlo, 2000). However, the difference between cognitive and propensity elements in CT does not indicate that they are mutually exclusive. Instead, they can be considered as two complementary dimensions. This could be exemplified by Ennis (2015) who proposed twelve dispositions and eighteen abilities connected to the nature of CT.

In addition to cognitive abilities and affective dispositions (Black, 2007), another aspect is CT styles. To illustrate, CT styles focus on how to demonstrate CT in problem-solving (Lamm, 2015). According to Lamm and Irani (2011), there are two CT styles: engagement and information-seeking. Engagers are more inclined to construct meanings from their surroundings and interactive communication. They are confident to demonstrate their thinking abilities in solving problems or drawing conclusions. On the other hand, seekers prefer to retrieve as much information as possible to raise their knowledge and find solutions to problems. Both styles are necessary, and ideal critical thinkers are expected to flexibly apply them in different contexts.

To obtain a comprehensive evaluation of CT performances of Chinese students, this paper defines critical thinking into three aspects: cognitive skills, affective dispositions and thinking styles (Baker, Lu, & Lamm, 2021; Ku, 2009).

### *2.2 Critical thinking of Chinese students*

Chinese students have been characterized as struggling in CT (Atkinson, 1997; Cortazzi & Jin, 1997; Ryan, 2010; Tian & Low, 2011). This seems to be the answer to the weak performance of Chinese students in academic writing and class discussion (e.g., Fakunle, Allison, & Fordyce, 2016; Guo & O'Sullivan, 2012). Research has been conducted to find out why this might be the case. Despite the ambiguity of its term (Byrne, 1994; Johnson, 1992), other elements such as cultural differences (Atkinson, 1997; Durkin, 2011), the unfamiliarity of western academic traditions (Paton, 2005; Turner, 2006), Chinese educational context (Lucas, 2019) and the poor proficiency in English (Floyd, 2011; Huang, 2008) may explain the poor CT performance among Chinese students.

Nevertheless, not all researchers seem to be satisfied with these claims. They indicate that this is a stereotype to regard Chinese students as poor critical thinkers (Lu & Singh, 2017; Tian & Low, 2011; Xu, 2021). To illustrate, the interpretation of CT in the Chinese context may deviate from that in western cultural traditions (Lu & Singh, 2017; Ryan, 2010). This is echoed by Heng (2018) who further clarified that this divergence may not necessarily imply the deficiency of Chinese students in CT. Unfortunately, some Chinese students may have accepted the alleged description of their deficit ability in CT (Song, 2014; Xu, 2021) which may negatively

influence their academic performance (Li, Chen, & Duanmu, 2010). Therefore, it is necessary to discover the real situation of Chinese learners' CT.

Indeed, several empirical studies have tried to answer this question (e.g., Liu, et al., 2018; Zhang & Lambert, 2008). However, some of them have methodological flaws, which would threaten the trustworthiness of their results (Gorard, 2013). Specifically, no standardised test is used in the judgement of CT (e.g., Fakunle et al., 2016; Guo & O'Sullivan, 2012; Li, 2013). While some researchers may show concern about the format of multiple-choice questions that involves a chance of guessing (Snyder, Edwards, & Sanders, 2019), the pre-specified evaluation criteria and the validation of testing items allow for a high level of objectivism. Another problem is that there is no comparison between Chinese students and their foreign peers (e.g., Ip et al., 2000; Zhang & Lambert, 2008). Even if Chinese students show positive results towards CT, it remains unknown whether they would perform better or worse in the international comparison. Irrespective of methodological issues, research has also been restricted in the higher education level (e.g., Loyalka et al., 2021; Yeh & Chen, 2003) and the nursing discipline (e.g., Yuan, Kunaviktikul, Klunklin, & Williams, 2008; Zhang & Lambert, 2008), which makes it difficult to generalize to the whole Chinese population.

## 3. Previous reviews in this area

There is a dearth of reviews that investigate the CT of Chinese students (e.g., Huang, 2019; Tian & Low, 2011), and those that did are focused on a broader group, such as Asian students in general (e.g., Indra, 2019; Salsali, Tajvidi & Ghiyasvandian, 2013). For example, Salsali et al. (2013) have compared the CT dispositions of Asian nursing students and those from other continents. Although they stated that a systematic method was used, there was no appraisal of the strength of evidence of the included studies. Hence, it is not possible to judge the evidence. Their review also included only peer-reviewed papers. This introduces publication bias (Song, Hooper, & Loke, 2013) since studies that report large, positive results are more likely to get published. These studies tend to be small-scale, using researcher-developed test instruments or do not include a comparator (i.e., single group, pre-post design). There is a large number of high-quality, large-scale, well-controlled studies that are unpublished. Cheung and Slavin's (2016) review found that 59% of these high-quality studies were unpublished. Excluding such high-quality studies can skew the results and lead to misleading conclusions (Slavin & Neitzel, 2020; Slavin, 2020a)

Among the very small minority of reviews that are concerned with only Chinese students, all were neither systematic nor critical. Tian and Low (2011), for example, provided critical insights on studies about Chinese students' CT dispositions. However, they did not search systematically, and therefore, no studies that consider the CT skill dimension was found. Huang's (2019) review was on high school students, but the study failed to evaluate the trustworthiness of the evidence. In other words, threats to validity, such as sampling strategy, sample size, attrition and conflict of interest were not considered.

This paper presents the results of a new systematic review of studies that compare the CT of Chinese students with other nationals to establish evidence for the common assumption about the lack of criticality of Chinese students.

## 4. Research aim and questions

The aim of this systematic review is to synthesise existing evidence on the CT of Chinese students. The primary objective of the study is to establish evidence for the common assumption about the lack of criticality of Chinese students. The achieve this objective, we frame our research question as: What is the reported performance of critical thinking of Chinese students, compared with students of other nationalities?

## 5. Methods used in the review

To address the research question, we review studies conducted that measure Chinese students' critical thinking. To establish the level of Chinese critical thinking, there needs to be a benchmark to judge the level by. Therefore, only studies that compare Chinese students' level of CT with that of other nationalities are included. A systematic review is, therefore, appropriate for the research question as it is comprehensive, transparent and systematic. In other words, it will help identify all relevant research relating to the research question. This ensures that the research that informs our conclusion is based on a comprehensive list of studies. This avoids cherry-picking only those studies that support the claim that Chinese students are lacking in criticality. Our systematic review also includes all published and unpublished reports (e.g., PhD theses), thus avoiding publication bias, where only research that agree with the popular conceptions or which align with the journals' or editors' stance are more likely to be accepted and published.

A systematic review permits evidence-based answers to research questions in a specific field through extensive searching, criteria-based selecting, critical evaluating, and unbiased analysing (Boland, Cherry, & Dickson, 2017; Klassen, Jadad, & Moher, 1998). Following a series of general stages such as identification, screening and including, this method explicitly delivers key information and increases the transparency of research (Boland et al., 2017; Hammersley, 2020). Besides, it allows for an in-depth analysis of existing literature (Siddaway, Wood, & Hedges, 2019), particularly when there are some disputes around a certain topic (Petticrew & Roberts, 2006). Since scholars hardly reached an agreement on the real situation of CT of Chinese students (e.g., Atkinson, 1997; Paton, 2005), it is appropriate to adopt the systematic review method in this research.

The review employed a protocol in line with current practice used in most systematic reviews. Broadly, it follows a series of stages as outlined in the Cochrane Review Handbook (Higgins et al., 2021). To ensure that the review is comprehensive, a systematic review approach was used to identify and evaluate existing studies, both published and unpublished.

### *5.1 Search strategy*

The first stage of the review is the development of the key search terms. These terms relate to the research questions, focused on "critical thinking" and "Chinese students". Accordingly, the keywords used in the search are:

("critical thinking" OR "think critically" OR "critical reasoning" OR "thinking skill*") AND (China OR Chinese) AND (student* OR learner* OR pupil*)

These terms are then applied and adjusted according to the idiosyncrasies of different databases.

*5.1.1 Databases*

Since the research topic is within the social science field, including education and psychology, we therefore searched for relevant studies in social science databases and search engines that host such databases. The EBSCO host search engine, for example, provides access to a range of databases, e-journals and e-books in education, psychology (e.g., PsycInfo) and social work. The databases are particularly useful for identifying journal articles and other publications on a particular topic within the subject areas covered by each database. Considering that the Education Resources Information Center (ERIC) hosts studies in education-related field (Boland et al., 2017), it therefore makes sense to also include ERIC in our search. To make sure that our search is comprehensive so that no relevant studies are missed, we also searched other databases including Applied Social Sciences Index & Abstracts (ASSIA), JSTOR, ProQuest, Sage Journals, Scopus, and Wiley online library. We included ProQuest as it covers Masters dissertations and PhD theses. This ensures that high quality unpublished work is also included in our review. We are, therefore, confident that our search is as comprehensive as it can be. This is where our review is different to previous reviews conducted on this topic.

The search was limited to studies from 2000 to 2021 as this was the period when the Chinese education system was reformed that emphasises CT. This period, therefore, saw an increase in research and publications on CT (Chen & Shi, 2017). Including studies from this period will also shed light on the impact of the education reform on Chinese students' CT capacity. The review included all published and unpublished materials. The search was also limited to studies published or reported in English or Chinese. The online database search was completed on 14 January 2022 and details are displayed in Appendix A.

*5.1.2 Manual searching*

To avoid publication bias, we also hand searched Google and Google Scholar to identify grey literature. In addition, references in the studies identified in the electronic database search were also followed up.

*5.2 Screening*

Relevant reports identified in the searches were then exported to EndNote (a reference managing software for screening). The first stage of screening was to remove duplicates, and to identify studies that are relevant to the research question. Prior to the screening, a list of inclusion and exclusion criteria was drawn. Studies were first screened for relevance by titles and abstracts by applying the inclusion and exclusion criteria. Then the full text was downloaded and screened.

*5.2.1 The inclusion criteria*

Studies were included if they were:

- Concerned with ethnic Chinese students (including students from Hong Kong, Taiwan, and Macau) and students from other nationals
- About students in schools or higher education
- Related to the assessment of critical thinking
- Empirical
- Published or reported between 2000 and 2021
- Published or reported in English or Chinese

*5.2.2 Exclusion criteria*

Studies were excluded if they:

- Focus solely on assessing the critical thinking of Chinese students with no comparison with other nationals
- Were not about students in schools or higher education (e.g., there were several studies that examined the critical thinking skills of individuals in different occupations; these were excluded.)
- Were theoretical pieces
- Were not primary research
- Did not have measurable outcomes of critical thinking (critical thinking skills, critical thinking disposition or critical thinking style)
- The outcomes were based on participants' self-report (i.e., subjective opinions or individual experiences)

For transparency, the screening process adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Page et al., 2021), which records the number of research reports found in database searches, the number included/excluded, the number screened, and the number retained and synthesised (Figure A).

*5.3 Data extraction*

The included studies were then data extracted where key information about each study's research design, sampling size, sampling strategy, outcome measures, missing data, method of analyses and the results was summarised (Appendix B). This information then informs the assessment of the strength of evidence. In this respect, the review is unique of reviews on this topic. Most previous reviews do not evaluate the trustworthiness of the findings by weighing the research evidence in terms of threats to validity.

*5.4 Quality assessment*

Quality assessment is crucial because if we are to have confidence in the findings of the review, the findings have to be based on the most robust evidence. To this end, each of the included studies is assessed for the trustworthiness of its finding using a quality appraisal tool, known as the "sieve" developed by Gorard (2021, p.94). The quality assessment is concerned primarily with the research design, the scale, threats to validity (e.g., attrition/missing data), and how outcomes are measured. The reputation of the authors and the publication outlets are ignored as each piece is judged solely on these criteria in the "sieve" (Table 1). To ensure inter-rater reliability, each study was rated by two reviewers. Where there was disagreement, a consensus was reached after discussion and careful review of the criteria.

**Table 1**
The Gorard "sieve" for quality assessment

| Design | Scale | Missing data | Measurement quality | Rating |
|---|---|---|---|---|
| Strong design for research question | Large number of cases (per comparison group) | Minimal missing data, no impact on findings | Standardised, independent, reasonably accurate | 4* |
| Good design for research question | Medium number of cases (per comparison group) | Some missing data, possible impact on findings | Standardised, independent, some errors | 3* |

| | | | | |
|---|---|---|---|---|
| Weak design for research question | Small number of cases (per comparison group) | Moderate missing data, likely impact on findings | Not standardised or independent, major possible errors | 2* |
| Very weak design for research question | Very small number of cases (per group) | High level of missing data, clear impact on findings | Weak measures, high level of error, or many outcomes | 1* |
| No consideration of design | A trivial scale of study | Hugh amount of missing data, or not reported | Very weak measures | 0* |

*5.5 Synthesis*

In the synthesis, the included studies were categorised according to the three dimensions of CT: critical thinking skills, critical thinking dispositions and critical thinking styles. Under each dimension, studies are grouped according to whether they report higher, lower or similar levels of CT (see Table 2 as an example). The strength of evidence for each level is determined by the number of studies and the quality rating. For example, if most of the studies rated 3* show mixed results, then we can safely say that the evidence for that dimension is mixed. The highest rated study is the one that informs the evidence. If none of the studies are rated above 2* and the number of studies are spread evenly across the levels, then it shows that the evidence is unclear. Similarly, if the majority of studies are rated 1*, and all of them show that Chinese students display lower CT, skills, we cannot conclude with confidence that Chinese students have lower CT skills as the evidence (demonstrated by the quality rating) is weak. The evidence is, therefore, only tentative.

## 6. Results

A total of 1,481 studies were retrieved from the online databases. Of these 735 were duplicates and thus removed. An additional 1,471 were identified from the manual search. Screening by title and abstracts removed 2,092 records that did not meet the inclusion and exclusion criteria. This retained 125 studies that were screened for full text. Of these, only 15 were deemed relevant to the research question and have met the inclusion and exclusion criteria (see Figure 1 for details).

No study was rated 4*, the highest rating possible. Only one study was assessed as 3*, and three studies as 2*. The remaining 11 studies were rated 1*. This indicates that the quality of research in the comparative analysis of Chinese students' CT vis-à-vis the CT of other nationals is generally poor.

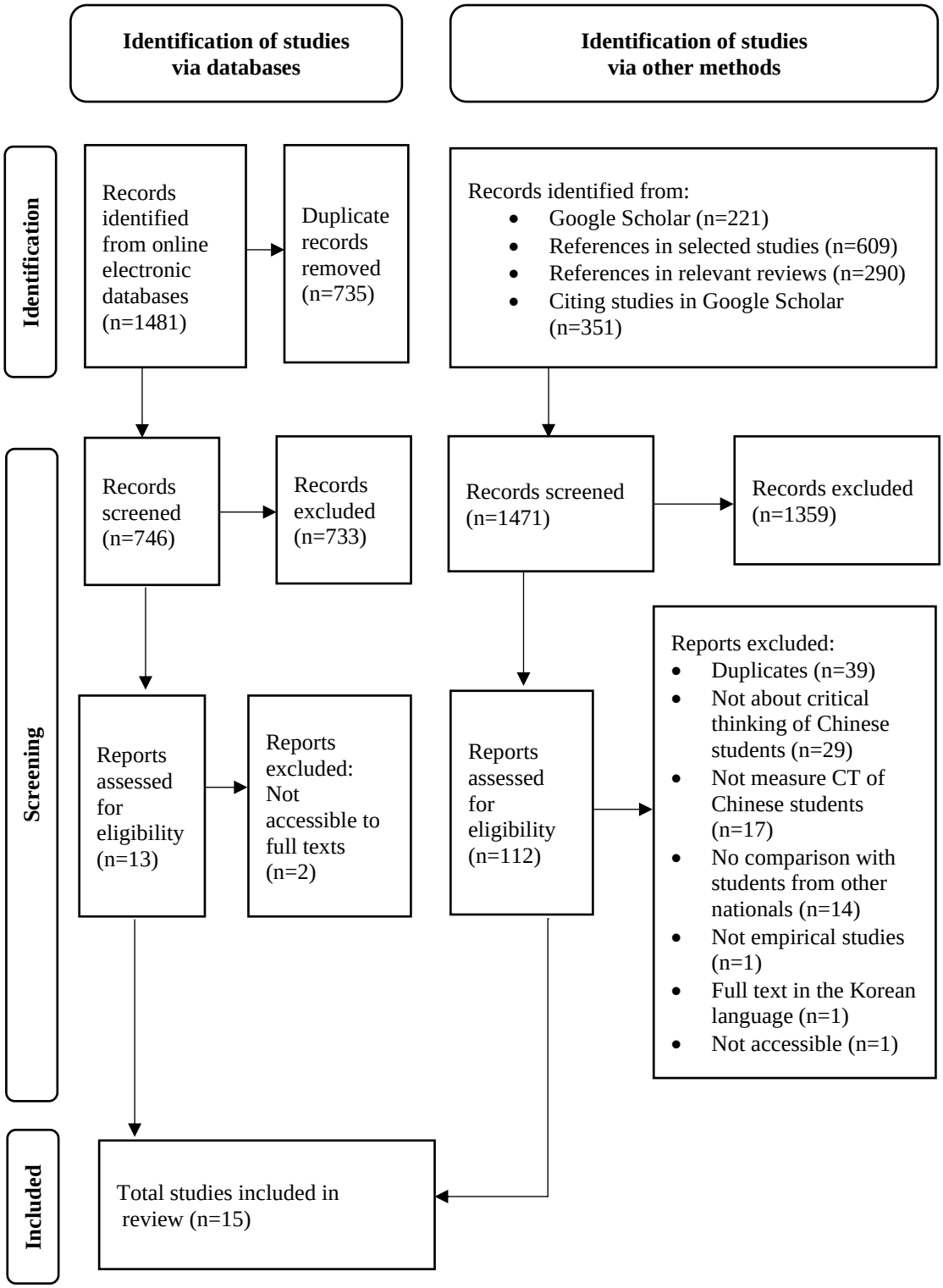


**Figure 1.** PRISMA flow diagram

*6.1 Discussion of findings*

*6.1.1 Critical thinking skills of Chinese students vs those of other nationals*

Of the 15 studies that met the inclusion criteria, nine compared CT skills of Chinese students and those of other nationals. CT skills include skills of interpretation, analysis, evaluation, inference, explanation and self-regulation, as well as deduction and assumption. Some studies

evaluated only a subset of these skills, while others included all these skills, depending on the test instruments used.

**Table 2**
Summary of comparison of critical thinking skills (n=9)

| Higher CT skills | Lower CT skills | No difference | Mixed | Strength of evidence |
|---|---|---|---|---|
| | | | | 4* |
| | | | 1 | 3* |
| 1 | | | 1 | 2* |
| 2 | 2 | | 2 | 1* |

Of the nine studies, three reported that Chinese students demonstrate higher CT skills than other nationals, two suggest that Chinese students had lower levels of CT skills, while four showed mixed results. The stronger studies (rated 2* or 3*) suggest mixed results (Hu, Adelopo, & Last, 2020; Loyalka et al., 2021) and one showed that Chinese students performed better than other nationals in CT skills test (Ku et al., 2006). This is at variance with the popular western stereotype perception of Chinese students as uncritical. The weaker studies are evenly distributed, with two reporting that Chinese students display higher CT skills, two suggesting that they have lower CT skills and two with mixed results. Therefore, based on the evidence presented, there is no evidence that Chinese students have higher or lower CT skills. The result is inconclusive.

**Loyalka et al. (2021)** compared the CT skills of Chinese, Indian, Russian and American students in two disciplines (computer science and electrical engineering). Chinese students' CT skills (measured by the HEIghten® suite of assessments from Educational Testing Service) showed that Chinese and American students had similar CT scores in the first two years, while Indian and Russian students had lower scores than Chinese and American students. In the fourth year, Chinese students performed similar to Russian students but higher than Indian students. However, compared to American students, Chinese students performed worse. While American students had improved in their CT skills in the last two years, all the other students showed a decline in the CT, but Chinese students showed the biggest decline compared to the other groups.

This is the largest study in this area involving more than 30,000 students across four countries. The study is rated 3* because of the large number of participants and careful consideration was given in the choice of instrument, languages and testing environment. The measurement instrument chosen from the study is designed to be culturally neutral. Similarly, to eliminate the influence of language, students were tested in their native language version of the CT test. Besides, the testing setting is natural, enhancing the degree of authenticity of scores.

Despite the care taken to ensure cultural comparability, there are important weaknesses in this study that lower the strength of the evidence to 3*. The number of participating students in each country is highly unequal. There were considerably more Indian students (n=17,455) and Chinese students (n=9,247) than Russian (n=4,703) and American students (n= 973). Although sampling weight was adopted to address the imbalance, it does not address potential biases in sample selection. According to their report, students in India, Russia and China were selected by random sampling, while students were volunteers in the US. Weighting for unequal sample size could multiply the bias particularly when American students were self-selected.

Besides, one of the problems is the high attrition rate among the US students (39%, calculated by the reviewer). While researchers acknowledged the missing data and addressed them by including missing value dummies in the regression, such replacement for missing data cannot usually overcome the bias introduced. Missing cases and missing data are seldom random (Gorard, 2020). Those that drop out of a trial or did not answer certain questions are likely to be different to those who did. Not considering missing cases is likely to overestimate the effect as missing cases are often non-random. For example, it is possible that those who drop out, or did not complete the test may be weaker students. Using weighting to overcome the missing cases among the US cohort may, in fact, magnify the bias. A further point is the restriction to only two disciplines (computer science and electrical engineering). There is also a gender imbalance in the sample. More than 60% of Chinese, Indian and Russian participants were males. This gender difference could help explain the difference between groups. According to Ennis, Millman, & Tomko (2005), gender is an important variable in measuring CT skills.

**Hu et al. (2020)**, a 2* study also showed mixed results. The study compared British and Chinese final year accounting and finance students in a British university. While Chinese students scored marginally higher than British students in inferential skills (55% vs 51%), they performed much worse than British students in tests of assumption, arguments and interpretation. On the test of deduction, Chinese students are on par with British students (62.5% vs 63%). The overall composite scores of Chinese students are lower than those of British students. This study is rated 2* because of the small number of cases (50 of each group). It is also not clear how the students were selected. Besides, a short version of the Watson-Glaser Critical Thinking Appraisal questionnaire (WGCTA) Form S was used. Further, it was translated into Chinese. This is particularly likely to introduce a possibility of error in translation. The process of testing is also problematic. Chinese students were initially tested using the English version test, and then the Chinese version. Both included the same content. This is likely to lead to familiarity with test items. Additionally, lecturers (who are not blinded), instead of researchers, administered the test which may introduce potential problems including inconsistency of research setting, unconscious bias (e.g., teachers may unconsciously give students greater support knowing that their scores will be compared). There is also the element of teacher expectation. All this can affect student performance in the test.

The study conducted by **Dong, Li and Liu (2010)** also suggests a mixed result. It compared Chinese students' CT skills with the norm of the four-year colleges and universities in the US. Specifically, Chinese students took the California Critical Thinking Skills Test (CCTST)-2000 designed by California Assessment Center (CAC) and their outcomes were compared with the norm data provided by the CAC. The results showed that the final-year Chinese undergraduates had an overall higher score in CT skills (mean 19.20, SD 4.32) than students in the US higher education (mean 16.80, SD 5.06). Although Chinese students scored higher than the norm in terms of the comprehensive CT ability, they performed lower in the skills of analysis and induction. Hence, this was a mixed result. The research is evaluated as 1* because of the very small number of Chinese cases (n=25), the majority of which were males (n=17).

**Liu (2013)** employed a similar research design and measurement instrument as Dong et al. (2010) but focused on Chinese second-year undergraduates (n=30) who were from two College English program classes at Xi'an Jiaotong University. Chinese students' overall CT skills scores (mean 19.83) are higher than those of American students whose scores were taken from the CAC (mean 16.80), but they have a lower level of inferential and inductive skills. This is in contrast to Hu et al.'s (2020) study, which reported that Chinese students performed better than British students on inferential skills. On other core skills, such as analysis, evaluation and

deduction, Chinese students outperformed their American counterparts. The students being compared may be different. In Hu et al.'s (2020) study students were final-year accounting and finance students, while those in Liu's (2013) study were second-year English programme students, most of whom were science majors from one top university in China. These students are, therefore, not representative of the average Chinese university students. It is also not clear if these students were compared with the general American undergraduate population, and whether the American and Chinese students were similar in terms of age and other demographic characteristics. Therefore, the study was rated 1* because this finding can only suggest a small advantage for Chinese students, but the results are far from conclusive given the lack of a similar comparator.

On the other hand, only one study rated 2* **(Ku et al., 2006)** indicates that Chinese students perform better than American students in CT assessments. Ku et al. (2006) recruited 142 Chinese students from a premier Hong Kong university and 153 American students from a public university in southern California. The Halpern Critical Thinking Assessment Using Everyday Situations (HCTAES) was adopted and translated to the Chinese language for the Chinese group. In terms of the overall scores, Chinese students (mean 119.20, SD 14.33) gained higher grades than U.S. students (mean 108.92, SD 18.11). However, five subscales including hypothesis testing, verbal reasoning, argument analysis, using likelihood and decision making/problem solving were not reported. Additionally, some background elements such as admission criteria and undergraduate major were not controlled. It is uncertain whether the two universities have similar levels of admission standards. Moreover, different majors may focus on different aspects of CT skills such as evaluation criteria (Bailin et al., 1999). While 77% of American participants majored in social science, only 40% of the Chinese cohort did. The disparity in majors is likely to influence the CT skills performance. Therefore, this study was rated as 2*.

**Zhang and Zhang (2013)** draw a similar conclusion in terms of the CT skills performance between Chinese and American learners. This study was rated 1* because oddly it used Pintrich, Smith, Garcia, and McKeachie's (1991) Motivated Strategies for Learning Questionnaire (MSLQ) to measure CT. In this study, 197 Chinese students from an English class and 165 American students from communication classes completed the test. To exclude the influence of language, the test was translated into Chinese (the alpha reliability 0.90). Their result suggested that Chinese students (mean 3.67, SD 0.92) perform better than U.S. students (mean 3.24, SD 0.87). However, the report did not explain how the samples were selected and how many did not respond. It is questionable whether comparing students in English and students in communication classes is a fair comparison as they may not be similar in terms of entry qualifications. Students may major in the two disciplines, or they attend these courses out of interest. It would be more helpful if more demographic information was included. In addition, MSLQ is an instrument designed to measure motivation and learning strategies rather than CT skills (Pintrich, Smith, Garcia, & Mckeachie, 1993). Moreover, the instrument was developed for use in western education systems (Zhang & Zhang, 2013), which may not be appropriate for Chinese learners. Another possible threat to the credibility of the result is that the American cohort was awarded extra credits for their participation, whereas Chinese students were not similarly incentivised. Hence, this study was 1*.

The study by **Park, Niu, Cheng, and Allen (2021)** is also 1*, reporting that Chinese students display higher CT skills. The purpose of this study was to investigate the cultural influence on CT in both Chinese (n=166) and American students (n=103). They extracted two vignettes from Lawson, Jordan-Fleming, and Bodle's (2015) Psychological Critical Thinking Exam, five

question items from CCTST and one vignette about an experimental generation from the Sternberg Scientific Inquiry and Reasoning. Both open-ended and close-ended questions were selected in their CT test. The final scores were averaged from these three assessments. This study found that Chinese students gained higher scores than their American counterparts in terms of CT skills (mean 1.32, SD 0.59 and mean 1.02, SD 0.44 respectively).

This study was rated 1* because the two groups compared were not equivalent. Some students had more advanced research experience while some had never been exposed to research. The research experience is positively correlated to CT skills (Haritania, Febrianib, Yulianac, & Arviana, 2019), indicating that those with more advanced research experiences may initially have better CT outcomes. As the selected samples were not randomised, the proportion of students with research experience and no experience in each group may be different. This may partially explain the difference in CT performance. Besides, their test only considered several key dimensions of CT, including evaluation, logical reasoning and probability thinking. Other aspects, such as analysis and deduction were not assessed. Using the average scores of the combined three tests may not be a good measure of general CT skills. For example, the test on scientific reasoning and enquiry may favour those with extensive research experience. Perhaps it is more informative to consider the weight of each test item. Furthermore, as Floyd (2011) suggested, Chinese students show lower scores when they take CT tests in English, compared to using their native language. It is also not clear if the tests were in English for both groups. If so, this might disadvantage the Chinese for whom English is not their first language.

**Lee et al. (2011)** found that compared to Korean nursing students (n=355), Chinese nursing students (n=407) demonstrate lower levels of CT skills (mean 94.43, SD 7.26 and mean 95.60, SD 8.59 respectively). The study was conducted in two Korean universities (four-year) and two Chinese universities (five-year). Although this study attempted to track changes in students' CT skills, it did not look at the same cohorts across years. What they did was to compare the first-year students with final-year students, and found that gains in CT scores between the first-year and final students were bigger for Korean students than for Chinese students. They then concluded that Korean students made a bigger improvement over time. But since this was not a longitudinal study, any difference between first-year and final-year students could simply be a reflection of the quality of students between cohorts. CT skills were measured using a critical thinking scale developed by Yoon (2004) and translated into Korean and Chinese versions. The translation to the respective language may have posed some problems. Besides, as mentioned by the authors, this is a self-reported questionnaire and may not be an accurate test of students' CT skills. The two groups being compared were also not equivalent. For example, the Korean freshmen were exposed to a course on CT whereas the Chinese students received no CT-related curriculum. The study, therefore, was rated 1*.

**Lun, Fischer and Ward (2010)** reported that Chinese students (n=24) displayed lower CT skills than their New Zealand European counterparts (n=35). The study included Asian students, with Chinese as a subset of these. The close-ended section of HCTAES was used as a measure of CT skills. Chinese students were found to perform worse (mean -1.26, SD 1.70) than New Zealand European students (mean 0.87, SD 1.13). However, the small number of non-representative participants from one university in New Zealand meant that the results cannot be to the wider Chinese student population, especially since the Chinese students were recruited from an international university rather than from local Chinese universities. Another issue is that all participants were tested in English. To what extent language may have impeded the performance of Chinese students, whose first language is not English is not known.

Additionally, the samples were asked to self-report their English proficiency, which is not a reliable measure of language proficiency. Therefore, the study was weak in evidence and rated 1*.

*6.1.2 Critical thinking dispositions*

According to Facione et al. (1995), CT disposition is about an internal tendency that leads to one's beliefs or actions. Some essential CT dispositional elements are truth-seeking, open-mindedness and inquisitiveness (see Ennis, 1985; Facione, 1990; Halpern, 1998).

**Table 3**

Summary of comparison of critical thinking dispositions (n=5)

| Higher CT disposition | Lower CT disposition | No difference | Mixed | Strength of evidence |
|---|---|---|---|---|
| | | | | 4* |
| | | | | 3* |
| | | | | 2* |
| | 3 | 1 | 1 | 1* |

The review identified five studies that measure and compare CT dispositions of Chinese students and those of other nationals (Table 3). Again, the result is somewhat mixed. Three studies showed that Chinese students had lower CT dispositions (McBride, Xiang, Wittenberg, & Shen, 2002; Tiwari, Avery, & Lai, 2003; Yeh & Chen, 2003), one showed no difference (Dennett, 2014) and one suggested a mixed result (Petrini & Kawashima, 2003). Although most of them suggest lower CT dispositions, this finding is not substantiated by stronger studies. The evidence is therefore inconclusive.

The study performed by **Yeh and Chen (2003)** compared the CT dispositions of a convenience sample of Taiwanese nursing students (n=214) with nursing students in the US (n=196). CT dispositions were measured using the translated version of the California Critical Thinking Dispositions Inventory (CCTDI). The study showed that Taiwanese students scored lower than American students on six subscales including truth-seeking, open-mindedness, analyticity, systematicity, self-confidence and maturity except for inquisitiveness. The mean overall score is 283.52 (SD 21.39) for Taiwanese students and 303.24 (SD 29.38) for American students (effect size = 0.8).

This study could have scored higher but because the sample was not randomised and it did not control for some background factors such as differences in age and working experience, it was rated 1*. Chinese learners were on average younger (mean age 22) while American students were older (mean age 28). Previous studies have shown that CT dispositions are correlated with age (Emir, 2009). Therefore, the difference in CT dispositions between Taiwanese students and American students may be the result of age rather than nationality. Previous research also suggests that nursing experience is positively correlated with CT dispositions (Feng, Chen, Chen, & Pai, 2010). And since almost half of the American students in this study had previous nursing experience (45.6%), while only 7.7% of Chinese did, previous experience may be an explanatory factor for the difference in CT dispositions between the groups. As the groups being compared are not similar in age, it is not possible to conclude either way. The disparity in age and work experience may explain the lower CT disposition scores of Chinese students. Another problem, as with other studies within the CT dispositions cohorts, is the self-report nature of CCTDI. While the tool is a standardised and independent measurement of CT dispositions (Facione et al.,1995), self-report is notoriously unreliable (Slavin, 2020b). The

convenience sampling also meant that participants may be self-selected. The non-random sample with unequal representation in each group and the self-report measures all reduced the reliability of this comparison.

Also focusing on the nursing students, **Tiwari et al. (2003)** investigated CT dispositions between Hong Kong Chinese students (n=222) and Australian learners (n=162). Their results indicated that Chinese students have lower CT dispositions (mean 268.36, SD 21.58) than their Australian counterparts (mean 287.73, SD 30.98). It is worth mentioning that Chinese nursing students also scored lower in all the seven subsets of CCTDI: truth-thinking, open-mindedness, analyticity, systematicity, self-confidence, inquisitiveness, and maturity. The research is rated as 1* because of a lack of control for age. Although the authors claimed that both Chinese and Australian students are similar in age, the age of Australians is not reported. Hence, as acknowledged by the authors, it remains unknown whether the age of Australian learners contributes to their CT dispositions performance. Another factor affecting the robustness of the results is the low response rate, with 61% for Chinese and 49% for Australian students. The low response rate potentially introduces non-response bias and may lead to misleading results (Prince, 2012).

**McBride et al. (2002)** compared the CT dispositions of Chinese (n=234) and American (n=218) physical education students using CCTDI. The researchers reported that Chinese undergraduates scored lower in truth-seeking, inquisitiveness, maturity, and self-confidence. However, scores in the other three aspects (analyticity, systematicity, and open-mindedness) were not reported because of the low Cronbach's alpha coefficients for Chinese samples. Perhaps the authors could have considered why these constructs had such low reliability instead of ignoring them. Although the number of students per cohort is comparable (234 vs 218), Chinese students were drawn from one Chinese university whereas the American students came from nine institutions. Any difference in CT disposition could be due to the kind of students in the one Chinese university and may not represent most Chinese students in higher education. There is also the issue of inconsistency in the data reported in the table and the text. For example, the table shows that Chinese students obtained a mean score in the maturity of 39.35, but in the text, it is reported that the maturity mean score was 30.35. The partial data, inconsistency in reporting and lack of fair comparison weaken the credibility of the findings. Therefore, 1* is given to this research.

While the studies above suggest that Chinese students have lower CT dispositions, **Dennett (2014)** found no difference between Chinese and American students in terms of CT dispositions. However, the evidence is weak for several reasons, such as the small non-random sample of students from one university. Only 41 Chinese and 50 American students participated in the research and all of them were from the same American university. Moreover, the English version of CCTDI was used for both groups. Since the language of CT assessments is evidenced to have an impact on students' performance (Floyd, 2011; Hu et al., 2020), it is inappropriate to use the CCTDI in English to evaluate Chinese students' CT dispositions. Any difference in performance could be attributed to language competency rather than CT. Comparing Chinese students studying in America with home students is not a fair comparison as Chinese international students who have chosen to study abroad are a biased group. They are likely to be more open-minded, are probably higher-performing students from well-to-do families. They are therefore not representative of Chinese students in general.

Things are more complex when Chinese students' CT dispositions are compared with those of learners from more than one country. While **Petrini and Kawashima (2003)** intended to

measure CT skills of Chinese, Japanese and Samoa nursing students, they used the CCTDI instrument, which measures CT disposition rather than skills. The researchers seemed to have confused dispositions with skills. For this reason, we include this study under CT dispositions rather than skills. The results show that Chinese students had higher CT dispositions (mean total score 277.75, SD 23.18) than Japanese students (mean total 271.84, SD 22.04). Although Chinese learners demonstrated a higher level of analyticity, systematicity, and self-confidence, they scored lower in truth-seeking, open-mindedness, inquisitiveness, and maturity. The results are therefore mixed. Comparisons of Chinese and Samoa learners showed no significant differences between the two groups. The overall results are mixed as there is no clear evidence that Chinese students have higher or lower CT dispositions compared to Japanese and Samoan students.

However, the evidence is not strong because of the convenience sampling with an unequal number in each group (165 Japanese, 300 Chinese and 70 Samoan). It is also unclear how students were recruited. While all the students in the three countries were females, they differed in terms of age and work experience. For instance, the Chinese students ranged in age from 21 to 25 and all of them had little clinical experience. Samoa students, on the other hand, ranged in age from 16 to 62, with extensive nursing experience. The failure to control these background elements casts doubts on the reliability of the findings. Considering the unequal number of cases in each country, the disparity in age and experience, the study was rated as 1*.

*6.1.3 Critical thinking styles: information seeking & engagement*

The critical thinking style focuses on the way an individual performs or expresses CT in practice (Lamm, 2015). Two kinds of CT styles have been identified and assessed: information seeking and engagement (Lamm & Irani, 2011). Information seekers acknowledge their limitations in knowledge or experience and are eager to gain more information before solving problems. Engagers show a desire to communicate and display confidence in explaining their reasoning process when making decisions. Lamm and Irani (2011) defined a good critical thinker as one who possesses both styles.

**Table 4**

Summary of comparison of critical thinking styles (n=1)

| Information seeking | Engagement | No difference | Mixed | Strength of evidence |
|---|---|---|---|---|
| | | | | 4* |
| | | | | 3* |
| 1 | | | | 2* |
| | | | | 1* |

Only one study **(Lu, Burris, Baker, Meyers, & Cummins, 2021)** that meets the inclusion criteria considers students' CT styles (Table 4). This study compared the CT styles of 104 U.S. students (37 males) and 103 Chinese students (69 males) majoring in agriculture. CT styles were measured using the University of Florida Critical Thinking Inventory (UFCTI), translated to Chinese for the Chinese version (Cronbach alpha 0.92). Only two constructs associated with CT styles were measured within the UFCTI: information seeking and engagement. Unlike instruments in the CT dispositions and skills, UFCTI measures students' preference for ways of CT expression and behaviours (Lamm & Irani, 2011). The study showed that American students scored higher on both engagement (mean 52.26, SD 6.25) and information seeking (mean 28.21, SD 3.55) than Chinese students (mean 45.97, SD 10.19) for engagement and mean 23.31, SD 5.30 for information seeking). Since the number of items measuring engagement and information seeking was not equal, the authors calculated the overall score by

transposing the seeking score and engagement score, and multiplied the engagement score by 1.866. Therefore, the mean overall score for Chinese students was 80.67 (SD 4.96) and 77.87 (SD 5.05) for the American students. Based on the UFCTI guidelines, students with an overall score above 79 are identified as seekers and those below 78 are engagers. The study suggests that Chinese students prefer information-seeking whereas American students are more inclined to an engaging CT style. While this does not tell us whether American or Chinese students have higher levels of CT, the different styles may help explain why Chinese students, on average, score lower on the CT skills test that measures analytical, evaluative, and deductive skills.

The study was assessed as 2 * due to some weaknesses. One is the lack of control over possible confounding factors. For example, most participants in the Chinese cohort were males whereas those from the US group were females. The gender difference may contribute to the difference in CT styles, and therefore, cannot be ignored. The other issue is the measurement quality. UFCTI requires students to self-report their styles, which is not an objective measurement. The CT styles of students may be related to their CT dispositions and skills. However, this review has found no studies that attempt to link these measures. Perhaps, future studies that attempt to compare CT skills of students could consider the relationship between CT styles and CT skills.

## 7. Conclusion

### *7.1 Limitations of the study*

As with any review of this scale, some relevant studies will have been missed, but the question is whether including these studies would have altered the results. Admittedly, the inclusion of English and Chinese language records published between 2000 and 2021 means that some potentially useful studies in earlier years may be missed. Besides, the systematic review is dependent on the existing literature. Although we aim to include all levels of education, most research still focuses on higher education. To the best of our knowledge, no study that compares CT performance of Chinese students and other learners has been conducted in primary, secondary, or high schools. This suggests a research gap in this area. Notably, the majority of studies on this topic are largely conducted in the nursing discipline. It is not clear why this is so, and why comparisons of Chinese students' CT with other nationalities are not more widely studied in other disciplines. Finally, although Chinese students have often been tagged as deficient in CT (Song, 2014; Xu, 2021), it is surprising that so few studies have actually tried to test if this is the case or not. Considering that CT includes multi-dimensions, however, no single study has explored CT skills, dispositions, and styles simultaneously.

### *7.2 Implications of the review*

The findings of the review suggest tentative evidence that there are differences in the CT dispositions, skills, and styles of Chinese students and those of other nationals. For example, the evidence suggests that Chinese students are less disposed to CT and more inclined to an information-seeking style, but the overall evidence is weak.

Most of the studies are conducted within one or two universities involving one cohort of students. Only one study considered the shift of CT across years (Loyalka et al., 2021). Most of the studies compared students in a particular discipline (e.g. nursing). Only one study evaluated students across disciplines and across more than one university in one country,

covering science, technology, engineering, and mathematics (STEM) education. Most of the studies reported substantial attrition or non-response. This is important as any missing cases can skew the results.

Ten of the 15 studies were conducted completely with undergraduate students. Studies involving postgraduate or doctoral students were rare. No studies conducted in the primary, secondary or high school sectors met the inclusion criteria. For example, while Chinese high school students' CT skills and dispositions were assessed, there was no comparison with foreign counterparts (Zhou, Wang, & Yao, 2007). Similarly, Fung's (2014) study only considered Hong Kong primary school students, with no comparison with other nationals.

Some studies compared Chinese students in China with American students in America (e.g., Ku et al., 2006; Lu et al., 2021), while others compared Chinese students with other students in American universities (e.g., Dennett, 2014). For example, Ku et al. (2006) compared American students in a Social Science faculty from a public university with Chinese students in a technical field from a premier university. The students being compared are not equivalent regardless of their nationality/ethnicity, therefore any differences in outcomes could be the result of differences in context and individual demographics. Most studies also did not control for differences in age and experience, nor did they consider the cultural contexts of the test. What these studies show is that measuring CT constructs across cultures is complex. As with some aspects of tests of intelligent quotient (IQ), these constructs may be culturally biased.

The mixed evidence prevents us from drawing a definitive conclusion. Among the included research, only one study involved random sampling (Loyalka et al., 2021). Future research needs to consider much larger samples across a range of schools or institutions. The groups compared should be equivalent in background demographics, courses of study and qualifications so that the major difference between groups is their nationality. Only thus can any differences in CT outcomes be attributed to country of origin. Studies need to use randomised controlled designs or equivalents, where observable and unobservable differences are controlled to ensure equivalence.

Additionally, the use of different CT instruments also introduces some biases. For example, research has suggested that second language proficiency could prevent students from demonstrating CT skills (Floyd, 2011; Manalo & Sheppard, 2016). It is justifiable to translate the CT tests into the Chinese language. However, some research fails to consider the influence of language (e.g., Dennett, 2014; Lun et al., 2010), and inappropriately uses western culturally based tests for Chinese cohorts. It is notable that all studies in the CT dispositions group chose CCTDI to measure the overall CT dispositions and seven subscales: truth-thinking, open-mindedness, analyticity, systematicity, self-confidence, inquisitiveness, and maturity. As acknowledged by the authors, the self-report measurement reduces the credibility of the tests (Paulhus & Vazire, 2007). It is not enough for future studies to choose a standardised and independent instrument. They need to consider the language and testing environment too.

There is no robust body of evidence indicating whether Chinese students have higher, lower or comparable levels of CT skills compared to other nationalities. The results vary depending on which groups of students were being compared. For example, Chinese students studying abroad may manifest a higher level of CT skills when compared with the home nation students. This could be because these students are different in terms of their prior attainment, their socio-economic background and perhaps academic ambition. Chinese students who can afford to study overseas or on scholarship tend to be from more privileged and higher socio-economic

families and are higher performers. The majority of the studies did not have equivalent comparison groups. None of the studies included in this review considered these factors or controlled for these confounding factors.

The glaring lack of good or even medium-quality studies suggests that this area of research is under-researched. It may be because many studies have uncritically accepted the belief that Chinese students have lower CT skills. This is the kind of image that some academics in western universities have of Chinese students – passive, unthinking and uncritical (Atkinson, 1997; Cortazzi & Jin, 1997; Zhang, 2017). This image may be perpetuated and have come to be accepted by many (even the Chinese themselves) as a true characterisation of Chinese students (Song, 2014). This widely accepted perception of Chinese students as docile, passive learners, lacking in criticality with no independent thoughts, is unhelpful and even damaging. It does not help students to develop their critical thinking especially if the assumption is that they are naturally uncritical. Students may come to accept that it is who they are, and resist any attempt to develop critical thinking. On the other hand, academics who accept this stereotypical perception of Chinese students may inadvertently be reinforcing this perception by treating them as passive, unthinking individuals. Bespoke curricula and pedagogy may be developed in western universities to support Chinese students in enhancing their critical thinking skills. This may be a wasted effort if Chinese students' passivity and reticence in voicing their disagreement are misinterpreted as a lack of criticality.

Moreover, several studies have shown that the lack of critical awareness and scepticism is not unique to Chinese students (See, 2016). For example, Arum and Roksa's (2011) study of over two thousand American students found that many university graduates do not know how to distinguish facts from opinion, or make clear written argument or objectively review conflicting reports. In England in the early 2000s when the Quality Assurance Agency (QAA) was introduced to monitor the quality of provision in higher education institutions, there was particularly concern that science students were unable to 'construct reasoned arguments on the ethical and social impact of advances in biosciences' (QAA, 2002). A report in the Independent newspaper (Independent, 2006, May 24) criticised the lack of argumentation skills among undergraduates in UK universities. Poets and authors called it a scandal that many of our supposedly brightest could not follow a logical train of thought or string a coherent argument. These examples illustrate that the perceived lack of criticality is not unique to Chinese students. The common misrepresentation of Chinese students is perhaps a misdiagnosis leading to wrong or inappropriate interventions. Our review suggests that there is no conclusive evidence that Chinese students are any less capable of critical analysis and argumentation when compared with other nationals.

See (2016) has long argued that critical thinking can and should be taught in schools and in higher education institutions. Argumentation skills should be integrated into content learning. If the existing instructional practices and modes of assessment in our schools (not only those in China) were revamped to require critical thinking, including analysis, synthesis, presenting opposing viewpoints, identifying logical fallacy and avoiding making assumptions, students will learn to develop these skills. Teachers and university lecturers are not just vehicles for the dissemination of knowledge, but educators to inspire and motivate young people to question, argue and critically evaluate information they receive. This is even more relevant now with the proliferation of fake news from social media.

**References**

Studies included in the structured review and ynthesis are marked with *


Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly, 31*(1), 71-94. doi:10.2307/3587975

Badger, J. (2019). A Case Study of Chinese Students' and IEP Faculty Perceptions of a Creativity and Critical Thinking Course. *Higher Education Studies, 9*(3), 34-44.

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies, 31*(3), 285–302.

Baker, M., Lu, P., & Lamm, A. J. (2021). Assessing the dimensional validity and reliability of the University of Florida Critical Thinking Inventory (UFCTI) in Chinese: a confirmatory factor analysis. *Journal of International Agricultural and Extension Education, 28*(3), 41- 56.

Barnett, R., & Davies, M. (2015). *The Palgrave handbook of critical thinking in higher education*. Basingstoke: Palgrave Macmillan.

Black, B. (2007). Critical Thinking-a tangible construct. *Research Matters: A Cambridge Assessment Publication, 2*, 2-4.

Black, B. (2012). An overview of a programme of research to support the assessment of critical thinking. *Thinking Skills and Creativity, 7,* 122-133. https://doi.org/ 10.1016/j.tsc.2012.04.003.

Byrne, M. (1994). *Learning to be critical*. Newcastle: Material and Resources Centre for Enterprising Teaching.

Chen, M., & Shi, N. (2017). Brief Review in Research Advance of Critical Thinking in China and Other Countries. *International Journal of Education, Culture and Society*, *2*(1), 13.

Cheung, A., & Slavin, R. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283-292.

Cortazzi, M., & Jin, L. (1997). Communication for learning across cultures. In D. McNamara & R. Harris (Eds.), *Overseas students in higher education: Issues in teaching and learning* (pp. 76–90). London: Routledge.

*Dennett, S. K. (2014). *A Study to Compare the Critical Thinking Dispositions between Chinese and American College Students.* Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED557684&site=ehost-live

*Dong, Y.X., Li, K, & Liu, F. (2010). The critical thinking skills of college English students: Assessment and cultivation. *Computer-assisted Foreign Language Educational in China, 135*, 33-38.

Durkin, K. (2011). Adapting to western norms of critical argumentation and debate. In J. Lixian & M. Cortazzi (Ed.), *Researching Chinese Learners: Skills, Perceptions and Intellectual Adaptations* (pp. 274–291). New York: Palgrave Macmillan.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity, 12*, 43-52. https://doi.org/10.1016/j.tsc.2013.12.004.

Emir, S. (2009). Education faculty students' critical thinking disposition according to achedemic achievement. *Procedia-Social and Behavioral Sciences, 1*(1), 2466-2469.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 44–48.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice.* 9-26. New York: W.H. Freeman and Company.

Ennis, R. H. (2015). Critical Thinking: A Streamlined Conception. In M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education.* 31-47. New York: Palgrave Macmillan.

Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell Critical Thinking Tests Levels X and Z Administration Manual.* Seaside, CA: Critical Thinking Company.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. The Delphi report: Research findings and recommendations. In *ERIC Doc. No. ED315-423*. ERIC Washington.

Facione, P. A., Facione, N. C., & Giancarlo, C. A. (2000). The Disposition Towards Critical Thinking: Its Character, Measurement and Relationship to Critical Thinking Skill. *Informal Logic. 20*(1): 61–84.

Facione, P. A., Sánchez, C. A., Facione, N. C., & Gainen, J. (1995). The disposition toward critical thinking. *The Journal of General Education, 44*(1), 1-25.

Fakunle, L., Allison, P., & Fordyce, K. (2016). Chinese postgraduate students' perspectives on developing critical thinking on a UK education masters. *Journal of Curriculum and Teaching, 5*(1), 27-38.

Feng, R. C., Chen, M. J., Chen, M. C., & Pai, Y. C. (2010). Critical thinking competence and disposition of clinical nurses in a medical center. *Journal of Nursing Research, 18*(2), 77-87.

Fisher, A. (2011). *Critical thinking: An introduction.* Cambridge university press.

Floyd, C. B. (2011). Critical thinking in a second language. *Higher Education Research and Development, 30*(3), 289-302.

Fung, D. (2014). Promoting critical thinking through effective group work: A teaching intervention for Hong Kong primary school students. *International Journal of Educational Research, 66*, 45-62.

Gorard, S. (2013). *Research design: creating robust approaches for the social sciences.* London: Sage.

Gorard, S. (2020). Handling missing data in numeric analyses. *International Journal of Social Research Methodology*, *23*(6), 651-660.

Gorard, S. (2021). *How to Make Sense of Statistics: Everything You Need to Know about Using Numbers in Social Science*. London: Sage.

Gorard, S., See, B. H., & Siddiqui, N. (2017). *The trials of evidence-based education: The promises, opportunities and problems of trials in education.* Routledge.

Guo, L., & O'Sullivan, M. (2012). From Laoshi to partners in learning: Pedagogic conversations across cultures in an international classroom. *Canadian Journal of Education, 35*(3), 164-179.

Halpern, D. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455.

Haritania, H., Febrianib, Y., Yulianac, T. P., & Arviana, E. (2019). The correlation of undergraduate course research experience and critical thinking skills. *International Journal of Innovation, Creativity and Change, 5*(6), 336-347.

Heng, T. T. (2018). Different is not deficient: Contradicting stereotypes of Chinese international students in US higher education. *Studies in higher education, 43*(1), 22-36.

Higgins, J. P. T, Thomas J, Chandler J, Cumpston M, Li, T, Page, M. J., Welch, V.A. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version

6.2 (updated February 2021). Cochrane, 2021. Retrieved from www.training.cochrane.org/handbook.

*Hu, L. Q., Adelopo, I., & Last, K. (2020). Understanding Students' Critical Thinking Ability: A Comparative Case of Chinese and British Undergraduates. *New Educational Review, 61*, 133-143.

Huang, R. (2008). Critical thinking: Discussion from Chinese postgraduate international students and their lecturers. *Hospitality, Leisure, Sport and Tourism Education, 4*(23), 1-12.

Huang, Y. (2019). Establishing Critical Thinking Course for High School Students in China: A Literature Review in Pedagogy Field. *Proceedings of the International Conference on Education, 5*(1), 59–66. https://doi.org/10.17501/24246700.2019.5107

Independent (2006). *University students: They Can't Write, Spell or Present an Argument.* Wednesday 24 May. http://www.independent.co.uk/news/education/higher/university-students-
they-cant-write-spell-or-present-an-argument-479536.html

Indra, V. (2019). Critical Thinking Disposition in Asian and Non-Asian Countries: A Review. *International Journal of Nursing Education and Research*, *7*(2), 279-282. 10.5958/2454-2660.2019.00063.2

Ip, W. Y., Lee, D. T., Lee, I. F., Chau, J. P., Wootton, Y. S., & Chang, A. M. (2000). Disposition towards critical thinking: a study of Chinese undergraduate nursing students. *Journal of Advanced Nursing*, *32*(1), 84-90.

Johnson, R. H. (1992). The problem of defining critical thinking. In S.P. Norris (Ed.), *The generalizability of critical thinking: Multiple perspectives on an educational ideal* (pp. 38-53). New York: Teacher College Press.

Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking skills and creativity, 4*(1), 70-76.

*Ku, K. Y. L., Chan, N.-M., Lun, V. M.-C., Haplern, D. F., Marin-Burkhart, L., Hau, K. T., & Ho, I. T. (2006, April). Chinese and United States undergraduates' critical thinking skills: Academic and dispositional predictors. Paper presented at the 2006 Annual Meeting of the American Educational Research Association: Education Research in the Public Interest, San Francisco, California.

Lamm, A. J. (2015). Integrating critical thinking into extension programming #3: Critical thinking style. Florida Cooperative Extension Service Electronic Data Information Source AEC546. https://edis.ifas.ufl.edu/wc208

Lamm, A. J., & Irani, T. (2011). *UFCTI manual*. Gainesville, FL: University of Florida.

Lawson, T. J., Jordan-Fleming, M. K., and Bodle, J. H. (2015). Measuring psychological critical thinking. Teach. *Psychol. 42*, 248–253. 10.1177/0098628315587624

*Lee, H. Y., Kim, Y., Kang, H., Fan, X. Z., Ling, M., Yuan, Q. H., & Lee, J. (2011). An international comparison of Korean and Chinese nursing students with nursing curricula and educational outcomes. *Nurse Education Today, 31*(5), 450-455. https://doi.org/10.1016/j.nedt.2010.09.002

Li, G., Chen, W., & Duanmu, J. L. (2010). Determinants of international students' academic performance: A comparison between Chinese and other international students. *Journal of studies in international education*, *14*(4), 389-405.

Li, Y. (2013). First year ESL students developing critical thinking: Challenging the stereotypes. *Journal of Education and Training Studies*, *1*(2), 186-196.

Lipman, M. (2003). *Thinking in education* (2nd ed.). Cambridge: Cambridge University Press.

*Liu, H. (2013). *Assessment of Students' Critical Thinking Skills in College English Program* 2012 international conference on education reform and management innovation, 1, 436-441.

Liu, O. L., Shaw, A., Gu, L., Li, G., Hu, S., Yu, N., Ma, L., Xu, C., Guo, F., Su, Q., Kardanovaj, E., Chirikov, I., Shi, J., Shi, Z., Wang, H., & Loyalka, P. (2018). Assessing College Critical Thinking: Preliminary Results from the Chinese HEIghten® Critical Thinking Assessment. *Higher Education Research and Development*, *37*(5), 999-1014.

*Loyalka, P., Liu, O. L., Li, G., Kardanova, E., Chirikov, I., Hu, S., Yu, N., Ma, L., Guo, F., Beteille, T., Tognatta, N., Gu, L., Ling, G., Federiakin, D., Wang, H., Khanna, S., Bhuradia, A., Shi, Z., & Li, Y. (2021). Skill levels and gains in university STEM education in China, India, Russia and the United States. *Nature Human Behaviour, 5*(7), 892-904. https://doi.org/http://dx.doi.org/10.1038/s41562-021-01062-3

*Lu, P., Burris, S., Baker, M., Meyers, C., & Cummins, G. (2021). Cultural Differences in Critical Thinking Style: A Comparison of US and Chinese Undergraduate Agricultural Students. *Journal of International Agricultural and Extension Education, 28*(4), 49-62.

Lu, S., & Singh, M. (2017). Debating the capabilities of "Chinese students" for thinking critically in anglophone universities. *Education Sciences, 7*, 1-16.

Lucas, K. J. (2019). Chinese graduate student understandings and struggles with critical thinking: A narrative-case study. *International Journal for the Scholarship of Teaching and Learning, 13*(1), 1-7.

*Lun, V. M. C., Fischer, R., & Ward, C. (2010). Exploring cultural differences in critical thinking: Is it about my thinking style or the language I speak?. *Learning and Individual differences, 20*(6), 604-616.

Manalo, E., & Sheppard, C. (2016). How might language affect critical thinking performance?. *Thinking Skills and Creativity*, *21*, 41-49.

*McBride, R.E., Xiang, P., Wittenberg, D. & Shen, J. (2002). An analysis of preservice teachers' dispositions toward critical thinking: A cross-cultural perspective. *Asia–Pacific Journal of Teacher Education, 30*(2), 131–140.

Moosavi, L. (2021). The myth of academic tolerance: the stigmatisation of East Asian students in Western higher education. *Asian Ethnicity*, 1-20.

Nilson, L. B. (2021). *Infusing Critical Thinking Into Your Course: A Concrete, Practical Approach*. Stylus Publishing, LLC.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hrobjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., et al., (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ 372*, n71. https://doi.org/10.1136/

*Park, J. H., Niu, W. H., Cheng, L., & Allen, H. (2021). Fostering Creativity and Critical Thinking in College: A Cross-Cultural Investigation. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.760351

Paton, M. (2005). Is critical analysis foreign to Chinese students? In E. Manalo & G. Wong-Toi (Eds.), *Communication skills in university education: The international dimension* (pp. 1–11). Auckland: Pearson Education.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality* (pp. 224 –239). London, England: Guilford.

*Petrini, M. A., & Kawashima, A. (2003). Comparison of Critical Thinking Skills of Nurses in Japan, China and Samoa. *Bulletin of the Graduate Schools Yamaguchi Prefectural University, 4*, 11-32.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide.* Oxford: Blackwell.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educational and Psychological Measurement, 53*(3), 801–813.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor, MI: The University of Michigan.

Prince, M. (2012). 9 - Epidemiology. In P. Wright, J. Stern, & M. Phelan (Eds.), *Core Psychiatry* (Third Edition) (pp. 115-129). W.B. Saunders. 10.1016/B978-0-7020-3397-1.00009-4

QAA (2002). *Quality Assurance Agency Subject Benchmark Statements: Biosciences*. Cheltenham: Quality Assurance Agency for higher Education; 2002.

Ryan, J. (2010). 'The Chinese learner': Misconceptions and realities. In J. Ryan & G. Slethaug (Eds.), *International education and the Chinese learner* (pp. 37–56). Hong Kong: Hong Kong University Press.

Salsali, M., Tajvidi, M., & Ghiyasvandian, S. (2013). Critical thinking dispositions of nursing students in Asian and non-Asian countries: A literature review. *Global Journal of Health Science*, *5*(6), 172-178.

See, B.H. (2016). An investigation into the teaching and learning of argumentation in first-year undergraduate courses: A pilot study. *British Journal of Education, Society and Behavioural Science*, 18 4, 1-25.

Slavin, R. & Neitzel, A. (2020). *In meta-analyses, weak inclusion standards lead to misleading conclusions. Here's proof.* [Blog post November 19]. Retrieved from https://robertslavinsblog.wordpress.com/category/published-vs-unpublished/ Accessed on 26 Feb 2022.

Slavin, R. (2020a). *Even Magic Johnson sometimes had bad games: Why research reviews should not be limited to published studies.* [Blog post February 27]. Retrieved from https://robertslavinsblog.wordpress.com/2020/02/27/even-magic-johnson-sometimes-had-bad-games-why-research-reviews-should-not-be-limited-to-published-studies/ Accessed on 26 Feb 2022.

Slavin, R. (2020b). *Reviewing Social and Emotional Learning for ESSA: MOOSES, not Parrots.* [Blog post May 25]. Retrieved from https://robertslavinsblog.wordpress.com/2017/05/25/reviewing-social-and-emotional-learning-for-essa-mooses-not-parrots/ Accessed on 5 March 2022.

Snyder, S. J., Edwards, L. C., & Sanders, A. L. (2019). An empirical model for infusing critical thinking into higher education. *Journal on Excellence in College Teaching, 30*(1), 127-156.

Song, F., Hooper, L., & Loke, Y. (2013). Publication bias: what is it? How do we measure it? How do we avoid it?. *Open Access Journal of Clinical Trials*, *2013*(5), 71-81.

Song, X. (2014). Changing social relations in higher education: the first-year international students and the 'Chinese learner' in Australia. In H. Brook, D. Fergie, M. Maeorg, & D. Mitchell (Eds.), *Universities in Transition: Foregrounding Social Contexts of Knowledge in the First Year Experience* (pp. 127-156). University of Adelaide Press.

Tian, J., & Low, G. D. (2011). Critical thinking and Chinese university students: A review of the evidence. *Language, Culture and Curriculum, 24*, 61-76. 10.1080/07908318.2010.546400

*Tiwari, A., Avery, A., & Lai, P. (2003). Critical thinking disposition of Hong Kong Chinese and Australian nursing students. *Journal of Advanced Nursing, 44*(3), 298-307.

Turner, Y. (2006). Students from mainland China and critical thinking in postgraduate business and management degrees: Teasing out tensions of culture, style and substance. *International Journal of Management Education, 5*(1), 3–11.

Watkins, D. A. & Biggs, J. B. (Eds.). (1996). *The Chinese learner: cultural, psychological, and contextual influences.* Hong Kong/Melbourne: CERC & ACER.

Xu, C. L. (2021). Portraying the 'Chinese international students': a review of English-language and Chinese-language literature on Chinese international students (2015–2020). *Asia Pacific Education Review*, 1-17.

*Yeh, M. L., & Chen, H. H. (2003). Comparison affective dispositions toward critical thinking across Chinese and American baccalaureate nursing students. *Journal of Nursing Research, 11*(1), 39-46.

Yoon, J. (2004). *Development of an instrument for the measurement of critical thinking disposition: In nursing*. Unpublished doctoral dissertation, The Catholic University of Korea, Seoul.

Yuan, H., Kunaviktikul, W., Klunklin, A., & Williams, B. A. (2008). Improvement of nursing students' critical thinking skills through problem-based learning in the People's Republic of China: A quasi-experimental study. *Nursing & Health Sciences*, *10*(1), 70-76. https://doi.org/10.1111/j.1442-2018.2007.00373.x

Zhang, H., & Lambert, V. (2008). Critical thinking dispositions and learning styles of baccalaureate nursing students from China. *Nursing & Health Sciences*, *10*(3), 175-181. https://doi.org/10.1111/j.1442-2018.2008.00393.x

*Zhang, Q., & Zhang, J. (2013). Instructors' positive emotions: Effects on student engagement and critical thinking in U S and Chinese classrooms. *Communication Education*, *62*(4), 395-411.

Zhang, T. (2017). Why do Chinese postgraduates struggle with critical thinking? Some clues from the higher education curriculum in China. *Journal of Further and Higher Education, 41*(6), 857-871. doi:10.1080/0309877X.2016.1206857

Zhou, Q., Wang, X., & Yao, L. (2007). A Preliminary Investigation into Critical Thinking of Urban Xi'an High School Students. *Frontiers of Education in China*, *2*(3), 447-468.

## Appendix A. Search syntax and results in databases

| Databases | Search syntax | Numbers of records |
|---|---|---|
| Applied Social Sciences Index & Abstracts (ASSIA) | ab ("critical thinking" OR "think critically" OR "critical reasoning" OR "thinking skill*") AND ab (China OR Chinese) AND ab (student* OR learner* OR pupil*) | 33 |
| EBSCO host<br>• Open dissertations<br>• British Education Index<br>• Education Abstracts<br>• ERIC<br>• APA PsycArticles<br>• APA PsycInfo | AB ("critical thinking" OR "think critically" OR "critical reasoning" OR "thinking skill*") AND AB (China OR Chinese) AND AB (student* OR learner* OR pupil*) | 280 |
| ProQuest:<br>• Dissertations & Theses Global<br>• ProQuest Social Sciences Premium | ab ("critical thinking" OR "think critically" OR "critical reasoning" OR "thinking skill*") AND ab (China OR Chinese) AND ab (student* OR learner* OR pupil*) | 497 |
| Sage Journals | [[Abstract "critical thinking"] OR [Abstract "think critically"] OR [Abstract "critical reasoning"] OR [Abstract "thinking skill*"]] AND [[Abstract China] OR [Abstract Chinese]] AND [[Abstract student*] OR [Abstract learner*] OR [Abstract pupil*]] | 20 |
| Scopus | (TITLE-ABS-KEY ("critical thinking" OR "think critically" OR "critical reasoning" OR "thinking skill*") AND TITLE-ABS-KEY (China OR Chinese) AND TITLE-ABS-KEY (student* OR learner* OR pupil*) AND PUBYEAR > 1999 AND PUBYEAR < 2022 | 373 |
| Web of Science | ab= ("critical thinking" OR "think critically" OR "critical reasoning" OR "thinking skill*") AND ab= (China OR Chinese) AND ab= (student* OR learner* OR pupil*） | 257 |
| Wiley online library | ""critical thinking" OR "think critically" OR "critical reasoning" OR "thinking skill*"" in Abstract and "China OR Chinese" in Abstract and "student* OR learner* OR pupil*" in Abstract | 21 |
| In total | 1481 | |

## Appendix B. Studies on critical thinking skills (n=9)

| Author(s) & date | Research design | Sample & level of education | Measuring instrument(s) | Finding(s) & result(s) | Limitation(s) | Rating |
|---|---|---|---|---|---|---|
| Loyalka et al. (2021) | A cross-sectional, cohort, comparative, and descriptive study | 5,102 freshmen and 4,145 junior Chinese students, 8,232 freshmen and 9,223 third-year Indian students, 2,607 freshmen and 2,096 third-year Russian students, and 973 undergraduate U.S. students<br><br>Sampling strategies in institutions: simple random sampling in China; stratified national random sampling in India and Russia; non-random sampling in the U.S.<br><br>Sampling strategies within the sample institutions: random sampling in China, India and Russia; non-random sampling in the U.S.<br><br>Undergraduate | Critical Thinking Exam, part of the HEIghten® suite of assessments from Educational Testing Service (ETS)<br><br>Translated to native languages in China, India and Russia | The freshmen and second-year Chinese students show similar critical thinking skills levels as their American counterparts, whereas their Indian and Russian peers are far lower.<br><br>Fourth-year Chinese university students demonstrate higher scores in critical thinking skills than Indian students, similar to Russian students, but much lower than the U.S. students in the fourth year.<br><br>Minimal gains in critical thinking skills are exhibited in the first two years in Chinese, Indian and Russian students.<br><br>Significant decrease in this aspect is evidenced in Chinese, Indian and Russian students during the last two years. On the contrary, American students show an increase in critical thinking skills during the final half of the university life.<br><br>**A mixed result** | Only focus on two disciplines (computer science and electrical engineering)<br><br>Not necessarily generalize to other contexts | **3*** |
| Hu, Adelopo, & Last (2020) | A cross-sectional study | 50 British students and 50 Chinese students<br><br>Not clear about sampling strategy<br><br>Final-year undergraduate | Watson-Glaser Critical Thinking Appraisal questionnaire (WGCTA) Form S<br><br>Modified: Content reduced to 20 questions in 5 sections (4 | Chinese students' inference skill score is higher than that of their counterparts (55% vs 51%).<br><br>However, scores of assumption, arguments and interpretation skills of Chinese students are lower than those of the English cohort, | Small scale study, restricted in only one UK university<br><br>A short duration of research time<br><br>Not a full WGCTA test | **2*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | questions per section)<br><br>Translated to a Chinese version | with 51% vs 72%, 41% vs 50%, and 58% vs 63% respectively.<br><br>The deduction skill scores between the two groups are similar, with 63% of English students and 62.5% of Chinese students.<br><br>Overall, Chinese students' critical thinking skills are poorer than that of British students.<br><br>**A mixed result** | | |
| Ku et al. (2006) | A correlational, cross-sectional study | 142 Chinese students (43 males, 99 females) and 153 U.S. students (30 males, 121 females, 2 with missing gender information)<br><br>Not clarify the sampling strategy<br><br>Undergraduate | Halpern Critical Thinking Assessment Using Everyday Situations (HCTAES)<br><br>Translated to Chinese language | Chinese students (mean 119.20, SD 14.33) gained higher scores than U.S. students (mean 108.92, SD 18.11) in terms of the critical thinking test.<br><br>**Higher CT skills** | The cross-sectional design | **2*** |
| Dong, Li, & Liu (2010) | A descriptive and comparative study | 25 Chinese undergraduates (8 females, 17 males)<br><br>Stratified random sampling<br><br>Final-year undergraduate | The California Critical Thinking Skills Test (CCTST)-2000 designed by California Assessment Center (CAC)<br><br>Translated to Chinese language | Chinese students' comprehensive critical thinking skills scores (mean 19.20, SD 4.32) are higher than those of American students (mean 16.80, SD 5.06).<br><br>Chinese students demonstrate a lower level in analysis (mean 3.52, SD 1.33 vs mean 4.44, SD 1.41) and induction (mean 9.32, SD 2.32 vs mean 9.53, SD 2.82), while higher in inference (mean 10.32, SD 2.40 vs mean 7.85, SD 2.69), evaluation (mean 5.36, SD 2.14 vs mean 4.52, SD 2.14) and deduction (mean | Small sample size, hard to be representative | **1*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 9.88, SD 2.68 vs mean 7.27, SD 2.89).<br><br>**A mixed result** | | |
| Lee et al. (2011) | A cross-sectional, comparative descriptive design | 355 Korean students and 407 Chinese students in nursing education<br><br>Stratified convenience sampling<br><br>All levels of undergraduate | Critical thinking Scale developed by Yoon (2004)<br><br>Translated to Korean language (Cronbach's alpha 0.85)<br><br>Translated to Chinese language (Cronbach's alpha 0.81) | Chinese students demonstrate lower scores of critical thinking (mean 94.43, SD 7.26), compared to Korean students (mean 95.60, SD 8.59).<br><br>**Lower CT skills** | Hard to control differences in nursing school systems, languages, and culture in these two countries<br><br>Self-reported questionnaires | **1*** |
| Liu (2013) | A descriptive study | 30 Chinese students majoring in sciences<br><br>Random sampling<br><br>Second-year undergraduate | The California Critical Thinking Skills Test (CCTST)-2000 designed by California Assessment Center (CAC)<br><br>Translated to Chinese language | Chinese students' overall critical thinking skills scores (mean 19.83, SD 2.74) are higher than those of American students (mean 16.80, SD not specified).<br><br>Chinese students demonstrate a lower level in inference (mean 7.37, SD 1.47) and induction skills (mean 7.18, SD 1.34), whereas other core skills including analysis (mean 4.93, SD 1.08), evaluation (mean 7.53, SD 1.72), deduction (mean 10.73, SD 1.91) are more proficient.<br><br>**A mixed result** | Not culturally neutral<br><br>Small sample size | **1*** |
| Lun, Fischer, & Ward (2010) | A comparison, correlational study<br><br>(Only consider the pilot study section because the main study includes a wider group: Asia students) | 24 Chinese students and 35 New Zealand European students<br><br>Not clarify the sampling strategy<br><br>In university level (not specify undergraduate, postgraduate, or other levels) | Halpern Critical Thinking Assessment Using Everyday Situations (HCTAES)<br><br>Only include the close-ended section of the HCTAES | Chinese students (mean -1.26, SD 1.70) perform worse than New Zealand European students (mean 0.87, SD 1.13) in the critical thinking test.<br><br>**Lower CT skills** | Only focus on the skill dimension of critical thinking<br><br>The paper-and-pencil form of assessment<br><br>Only use one test to measure critical thinking | **1*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| Park, Niu, Cheng, & Allen (2021) | A correlational and cross-sectional study | 166 Chinese and 103 American students<br><br>The internet-based contact method (not specify the sampling strategy)<br><br>In university level (not specify undergraduate, postgraduate, or other levels) | An updated Psychological Critical Thinking (PCT) Exam by Lawson et al. (2015)<br><br>California Critical Thinking (CCT) Skills Test<br><br>The experimental generation part from Sternberg Scientific Inquiry and Reasoning<br><br>Averaged scores of these three tests: experiment generation (one vignette), PCT (two vignettes), and CCT (five sample items) | Chinese students (mean 1.32, SD 0.59) outperform American students (mean 1.02, SD 0.44) on critical thinking.<br><br>**Higher CT skills** | Low level of representativeness of participants due to gender and discipline differences<br><br>Only focus on three dimensions: evaluation, logical reasoning and probability thinking | **1*** |
| Zhang & Zhang (2013) | A correlational, cross-sectional study | 197 Chinese students and 165 U.S. students<br><br>The class-based contact method (not specify the sampling strategy)<br><br>In university level (not specify undergraduate, postgraduate, or other levels) | Motivated strategies for learning questionnaire (MSLQ) from Pintrich et al (1991)<br><br>Adopt the critical thinking subscale (the alpha reliability for U.S. 0.86)<br><br>Translated to Chinese (the alpha reliability for Chinese 0.90) | Chinese students (mean 3.67, SD 0.92) perform better than U.S. students (mean 3.24, SD 0.87) in the critical thinking test.<br><br>**Higher CT skills** | The instrument characteristics: developed in the U.S., likely to be inappropriate for Chinese students<br><br>Self-report responses | **1*** |

**Studies on critical thinking dispositions (n=5)**

| Author(s) & date | Research design | Sample & level of education | Measuring instrument(s) | Finding(s) & result(s) | Limitation(s) | Rating |
|---|---|---|---|---|---|---|
| Dennett (2014) | A cross-sectional, comparative study | 41 Chinese and 50 American students<br><br>Voluntary sampling<br><br>In both undergraduate and postgraduate levels | California Critical Thinking Disposition Inventory (CCTDI) | No significant differences in critical thinking dispositions are identified between Chinese and American students.<br><br>**No difference** | Difficult to generalize because of the voluntary sampling<br><br>Bias introduced by the researcher's experience of teaching<br><br>Use only one instrument to measure critical thinking<br><br>Closed-ended instrument, no space for alternatives<br><br>Need to consider factors such as Chinese students' choice of studying abroad, prior experiences and university teachers' methods to develop critical thinking | **1*** |
| McBride, Xiang, Wittenberg, & Shen (2002) | A cross-cultural, comparative, and descriptive study | 218 American students and 234 Chinese students in physical education programmes<br><br>Selective sampling for American universities, and voluntary sampling for American students; purposive sampling for Chinese students<br><br>Undergraduate: juniors or seniors | The California Critical Thinking Dispositions Inventory (CCTDI)<br><br>Translated to Chinese language (reliability coefficient 0.78) | American students score higher in truth-seeking [mean 35.17 (from the table) /38.17 (from the text), SD 5.59 vs mean 34.62, SD 5.65], inquisitiveness (mean 44.01, SD 8.91 vs mean 43.29, SD 5.80), maturity [mean 42.66, SD 6.75 vs mean 39.35 (from the table)/ 30.35 (from the text), SD 6.08] and self-confidence (mean 43.90, SD 6.69 vs mean 40.72, SD 6.02) than Chinese students. | Hard to generalize because of the Chinese sampling strategy | **1*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | **Lower CT dispositions** | | |
| Petrini & Kawashima (2003) | A cross-sectional, comparative, descriptive study | 165 Japanese (82 students are 21-25 years old with no nursing related experiences; 83 students are with at least 5 years of experience), 300 Chinese (all are 21-25 years old and hardly have clinical experience) and 70 Samoa nursing students (all are 16-62 years old and with diverse nursing experience)<br>Convenience sampling in each country<br><br>Undergraduate | The California Critical Thinking Dispositions Inventory (CCTDI)<br><br>Translated to Japanese (Cronbach's alpha 0.83) and Chinese languages (Cronbach's alpha 0.81) | A significant difference in critical thinking is evidenced between Japanese and Chinese students (Tukey's Honestly Significant Difference: $P<0.05$). However, there is no difference between Chinese and Samoa students ($P>0.05$, Tukey's Honestly Significant Difference: not specified).<br><br>The total scores of CCTDI of Chinese students (mean 277.75, SD 23.18) are higher than Japanese (mean 271.84, SD 22.04).<br><br>Chinese students show lower scores in truth-seeking (mean 31.38, SD 5.32 vs mean 34.87, SD 5.17), open-mindedness (mean 37.52, SD 4.73 vs mean 41.78, SD 4.15), inquisitiveness (mean 46.28, SD 5.77 vs mean 46.64, SD 5.48), and maturity (mean 36.93, SD 6.51 vs mean 43.73, SD 5.21), while higher in analyticity (mean 42.34, SD 5.38 vs mean 36.59, SD 4.48), systematicity (mean 38.84, SD 5.05 vs mean 35.13, SD 5.48) and self-confidence (mean 44.47, SD 6.04 vs mean 33.10, SD 7.51), compared with the Japanese cohort. | Small sample size and convenience sampling, hard to generalise results<br><br>Lack of some demographic information (e.g., educational background, admission criteria)<br><br>The cultural-embedded instrument | **1*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | **A mixed result** | | |
| Tiwari, Avery, & Lai (2003) | A cross-sectional, descriptive, and comparative study | 222 Hong Kong Chinese students and 162 Australian nursing students<br><br>Convenience sampling<br><br>All levels throughout the pre-registration and post-registration nursing programme | The California Critical Thinking Disposition Inventory (CCTDI)<br><br>Translated to Chinese language (Overall alpha 0.70) | Chinese students scored lower in all seven aspects: truth-seeking (mean 31.30, SD 4.52 vs mean 35.03, SD 6.94), open-mindedness (mean 38.40, SD 3.70 vs mean 41.86, SD 6.22), analyticity (mean 41.32, SD 4.12 vs mean 41.73, SD 6.01), systematicity (mean 37.13, SD 4.97 vs mean 38.51, SD 6.16), self-confidence (mean 40.27, SD 5.83 vs mean 40.74, SD 6.50), inquisitiveness (mean 43.60, SD 5.79 vs mean 46.29, SD 6.56), and maturity (mean 36.34, SD 5.29 vs mean 43.57, SD 6.74).<br><br>Overall, Chinese students display a negative critical thinking disposition (mean 268.36, SD 21.58), whereas the Australian group are more inclined to positive ones (mean 287.73, SD 30.98).<br><br>**Lower CT dispositions** | Hard to generalize results because of the snapshot design, convenience sampling and high level of missing data | **1*** |
| Yeh & Chen (2003) | A comparative, correlational, cross-sectional research design | 214 nursing Chinese students in Taiwan and 196 nursing students in the USA<br><br>Convenience sampling<br><br>Undergraduate (juniors and seniors) | California Critical Thinking Dispositions Inventory (CCTDI)<br><br>Translated to Chinese language (overall Cronbach's alphas 0.71) | Chinese students gain lower scores in six subscales including truth-seeking (mean 30.97, SD 4.86 vs mean 39.15, SD 6.29), open-mindedness (mean 40.90, SD 4.60 vs mean 43.90, SD 5.70), analyticity (mean 43.01, SD 4.09 vs mean 43.06, SD 5.50), | Self-report critical thinking dispositions<br><br>Convenience sampling<br><br>Low level of generalisability due to the cross-sectional design<br><br>Use different language | **1*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | systematicity (mean 38.28, SD 5.17 vs mean 41.11, SD 6.60), self-confidence (mean 42.47, SD 6.14 vs mean 42.94, SD 6.67) and maturity (mean 39.47, SD 5.14 vs mean 45.73, SD 6.96) except for the inquisitiveness (mean 48.42, SD 5.39 vs mean 47.34, SD 6.35).<br><br>Overall, Chinese students show lower scores in critical thinking dispositions (mean 283.52, SD 21.39) than American undergraduates (mean 303.24, SD 29.38).<br><br>**Lower CT dispositions** | versions of CCTDI | |

**The study on critical thinking styles (n=1)**

| Author(s) & date | Research design | Sample & level of education | Measuring instrument(s) | Finding(s) & result(s) | Limitation(s) | Rating |
|---|---|---|---|---|---|---|
| Lu, Burris, Baker, Meyers, & Cummins (2021) | A cross-sectional study | 104 U.S. students (37 males) and 103 (69 males) Chinese students majoring in agriculture<br><br>Convenience sampling<br><br>Undergraduate | University of Florida Critical Thinking Inventory (UFCTI)<br><br>Translated to a Chinese version (Overall reliability measured by the Cronbach's alpha 0.92) | Chinese students scored lower in engagement (mean 45.97, SD 10.19) than American students (mean 52.26, SD 6.25).<br><br>Chinese students also scored lower in information seeking (mean 23.31, SD 5.30) than American students (mean 28.21, SD 3.55).<br><br>U.S. students are more inclined to an engaging critical thinking style (mean 77.87, SD 5.05), whereas Chinese students prefer an information-seeking critical thinking style (mean 80.67, SD 4.96). [The overall scores are transposed and multiplied the engagement score by 1.866 due to the unequal number of items.]<br><br>**Information seeking** | Using a convenience sample, limited in one university in each country, low level of generalizability<br><br>Only exploring two constructs within critical thinking styles<br><br>Only use one variable (country) to measure cultural differences | **2*** |