Can the Infusion Teaching of Critical Thinking Improve Chinese Secondary Students' Critical Thinking and Academic Attainment? Findings from a Randomised Controlled Trial

Keji Fan[1,*]

[1.] School of Education, Durham University, Durham, DH1 3LE, United Kingdom

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

Abstract

Critical thinking has been recognised as a core skill in the latest English curriculum standards for secondary schools in China. However, the current state of critical thinking education in China remains underdeveloped. Although some Chinese secondary schools have reported success in infusing critical thinking into regular teaching, there is little evidence of its impact as no independent robust studies have been conducted. The aim of this study was to determine whether the infusion of critical thinking in the regular curriculum has any impact on improving Chinese secondary students' critical thinking skills and academic attainment using a more robust research design that can establish causation. Twenty-one English language teachers and 2,011 Grade 8 students from four rural secondary schools in China participated. The intervention was delivered by 11 experimental English language teachers once a week for three months. The impact of the intervention was measured by differences in modified standardised critical thinking skill tests. A process evaluation, including class observations and interviews, was also conducted to determine fidelity to the treatment. The results indicate a small positive impact of the infusion teaching on students' critical thinking skills. Therefore, it is feasible and promising to infuse critical thinking into the English curriculum in China's secondary schools. However, the positive impact was not observed in academic attainment outcomes. The trial provides policy and educational practice recommendations to support the development of critical thinking skills among Chinese students.

Keywords: infusion approaches; critical thinking; academic attainment; Chinese secondary students

1. Background

With the rapid proliferation of information and its easy accessibility through social media platforms, students are continually bombarded with various types of information. It is important that students are equipped with the critical thinking (CT) skills to distinguish between accurate information and misinformation (Horn & Veermans, 2019; Nilson, 2021). Research indicates that even though advertisements are clearly labelled as sponsored content on social media platforms, most readers still struggle to differentiate genuine news stories from ads and unauthoritative posts (Tsipursky & Ward, 2020). The hidden algorithms used to filter information exacerbate the problem and promote the illusion of a false consensus on contentious issues (Nilson, 2021). In the current climate, what students need is not more information, but CT skills, the ability to evaluate evidence and discern what is believable and what is not. Such skills are particularly essential for young people in the 21st century (Organisation for Economic Co-operation and Development [OECD], 2018).

Over the last three decades, policymakers in China have increasingly recognised the importance of educating citizens to think critically, creatively, and innovatively to compete in the global economic and technological arena (OECD, 2011). They issued a report *"Outline of China's National Plan for Medium and Long-term Education Reform and Development (2010-2020)"* (People's Republic of China, 2010) that focused on cultivating well-rounded citizens and improving the education system. Although it recognised the importance of learning with thinking, it provided insufficient details on which CT skills were desired and how to enhance them. Until April 2022, the Ministry of Education (MoE) issued a new curriculum plan and standards for primary and secondary schools, with a clear emphasis on CT (MoE, 2022a, 2022b). These have been implemented since September 2022. The new English curriculum standards for secondary schools (MoE, 2022b) emphasise skills such as interpretation, analysis, evaluation, inference, explanation, deduction, assumption, and self-regulation, all of which are elements of CT skills.

Despite policy efforts to foster CT in schools, there are many practical challenges in China. Firstly, the introduction of CT is primarily implemented in the higher education sector, with students in primary and secondary schools rarely exposed to CT teaching (Dong, 2015). Indeed, some schools in China have explored ways to embed CT in the regular curriculum (Li, 2017; Zhai, 2015). One method is the infusion of CT into the regular curriculum, where subject content is used, and the improvement of CT is made explicit to students (Ennis, 1989). The infusion method is deemed appropriate in the Chinese context and is more likely to be accepted by schools as it is integrated into the existing curriculum using textbooks and materials already in use, rather than as a standalone course (Tan, 2020). It also fits well with school timetabling. Despite these Chinese schools' claimed success in the implementation of the infusion method, the teaching of CT in Chinese secondary schools remains ad hoc, sporadic, unstructured, and unsystematic. Classroom pedagogies have largely remained unchanged (Zhao, 2020). The extent to which CT is emphasised varies between schools and regions. Many schools, especially those in rural areas, still rely heavily on traditional teacher-centred approaches. Classroom teaching remains didactic, with passing exams and rote learning still dominating the Chinese schooling system (Dello-Iacovo, 2009; Zhao, 2020). Perhaps due to this, there are still mixed results on Chinese students' CT skills compared to those of other nationalities (Fan & See, 2022). As this trial is conducted in secondary rural schools in China, it enriches the current CT education in China where urban schools and higher education institutions have primarily focused on CT teaching.

Secondly, Chinese educators are faced with challenges in fostering CT skills amidst high-stakes assessments and limited teacher training. Teachers in China have not been adequately prepared to teach CT. They continue to face pressure to prepare students for high-stakes assessments (Tan, 2020). Most teacher training has focused solely on the theory and importance of CT, lacking sufficient high-quality training with hands-on activities to enable them to effectively teach CT in the classroom (Yan, 2012; Zou & Lee, 2023). As part of the research, a suite of teaching and learning resources will be developed for use in schools, and teachers will receive training on how to embed CT teaching into the existing curriculum.

Most current research on CT education in China is conducted in the higher education sector (e.g., Cheng, Huang, Yang, & Chang, 2020; Cui & Teo, 2023; Dong, 2017; Yuan, Kunaviktikul, Klunklin, & Williams, 2008; Zhang et al., 2017). To our knowledge, there are currently few studies that are focused on secondary school students in China (e.g., Fung, 2017; Ku, Ho, Hau, & Lai, 2014; Wang, Chen, Lin, Huang, & Hong, 2017). Additionally, many previous studies on the impact of the infusion method have methodological limitations. On the one hand, the small sample size limits the generalisability of their findings. Many studies recruited no more than 100 students (e.g., Bağ & Gürsoy, 2021; Dong, 2017; Lin, 2014; Marin & Halpern, 2011; Zohar & Tamir, 1993). On the other hand, many studies do not carefully consider the fidelity to treatment and diffusion problems (Toomey et al., 2020). For instance, in the quasi-experimental study conducted by Bağ and Gürsoy (2021), it is unknown whether the instructor taught both experimental and control classes. If this is the case, there might be a diffusion problem where the teacher unconsciously taught CT-relevant content to the control class. If there were two teachers, any impact might be due to teacher differences and cannot be exclusively attributed to the intervention. Similarly, in Hu et al.'s (2011) study, experimental and control students were in the same class, and in Lin's (2014) study, the same teacher taught both the infusion and traditional English courses. This considerably increased the problem of diffusion, but neither study made any efforts to prevent it.

Overall, there have been no robust evaluations of the infusion approach on Chinese secondary students' CT skills and academic attainment. To fill the research gaps, this study employs a randomised controlled trial with a substantial sample size of over 2,000 Chinese secondary students and 21 teachers and involves a process evaluation. This study holds significance in offering a practical way of introducing CT teaching into schools with minimal disruption to the existing curriculum, developing students' CT competency and addressing the Chinese government's ambition for its citizens to remain competitive in the global education landscape.

2. Literature review

2.1 What is critical thinking?

Critical thinking is a contentious term (Byrne, 1994; Fisher, 2011; Nilson, 2021) and draws from fields such as philosophy, cognitive psychology, and education (Lai, 2011). Each discipline offers its own interpretation of what constitutes CT. Philosophically, CT is considered a form of rational and self-regulatory thinking that individuals use to justify their decisions or actions (Ennis, 1987). However, this definition tends to exaggerate the outcome of thinking and neglects the specific thinking processes (Lipman, 2003) or elements of reasoning (Elder & Paul, 2020). Cognitive psychologists, on the other hand, argue that the philosophical definitions of CT are too idealistic, assuming how people would behave under the "best" conditions (Sternberg, 1986). They have attempted to frame CT from a different perspective, considering its nature in authentic contexts (Black, 2007). According to psychologists, CT is a construct that cannot be directly observed (Bailin, 2002). To define CT in a way that can be observed and measured, cognitive psychologists focus on the products of CT, including a set of behaviours and skills exhibited by critical thinkers (Halpern, 1999; Lai, 2011). Willingham (2007), for instance, listed a collection of critical thinkers' behaviours: "seeing both sides of an issue, being open to new evidence that disconfirms your ideas, reasoning dispassionately, demanding that claims be backed by evidence, deducing and inferring conclusions from available facts, solving problems, and so forth" (p. 8). Furthermore, educational psychologists focus on learning and instruction, including how CT can be developed (Barnett & Davies, 2015). Bloom's taxonomy is sometimes recognised as CT, involving skills such as memorising, understanding, applying, analysing, evaluating, and creating (Krathwohl, 2002). From the educational perspective, CT is seen as a desired goal that can be taught, measured, and applied (Dhakal, Watson Todd, & Jaturapitakkul, 2023).

It is clear that the definition of CT is complex and varies across disciplines (Black, 2007). In this study, CT skills are the focus as they are teachable, measurable, and applicable, which is appropriate and tangible in the education field (Dhakal et al., 2023).

Defining CT skills is also complex. Multiple models of CT skills have been proposed, each offering unique perspectives and frameworks to understand and cultivate CT skills. Ennis (2015), a leading figure in this field, suggested fifteen skills associated with an ideal critical thinker who can draw well-founded conclusions. These can be summarised as analysing arguments, evaluating sources, making inferences and judgements, and clarifying concepts. Additionally, using the Delphi method, Facione (1990) found a consensus among a panel of 46 scholars from various disciplines that

CT involves a set of six cognitive skills: interpretation, analysis, evaluation, inference, explanation, and self-regulation.

Paul and Elder (2010) also made significant contributions to the field. They believed that CT entails independent, disciplined, self-monitored, and self-adjusting thought processes, requiring trustworthy standards to analyse and evaluate the structures inherent in thinking. According to their framework, human thinking is naturally biased and distorted (Elder & Paul, 2020), thus necessitating systematic cultivation of CT. This requires rigorous intellectual standards, including *clarity, accuracy, precision, relevance, depth, breadth, logic, significance, and fairness.* They propose five key CT skills:

- Asking clear and precise questions
- Collecting and assessing relevant information, and interpreting it effectively
- Drawing reasonable and logical conclusions or solutions
- Considering alternative explanations, identifying and evaluating assumptions
- Communicating effectively with others to collaboratively devise solutions to complex problems

Despite conceptual variations in the literature regarding CT skills, some common features emerge. Firstly, there is a general agreement that CT skills include rigorous analysis, evaluation of arguments, and logical reasoning. Secondly, CT is seen as a composite cognitive skill comprising several sub-skills. These sub-skills, while independent, are closely interrelated (Bailin, Case, Coombs, & Daniels, 1999; Dhakal et al., 2023). Aligning with these commonalities, this study considers CT as a composite of multiple sub-skills, focusing on analysis, evaluation, and inference.

2.2 The infusion approach

Ennis (1989) proposed four approaches to CT teaching: general, infusion, immersion, and mixed. In the general approach, CT is taught independently of specific subject matter, with CT objectives explicitly outlined to students. Both curriculum content and CT improvement are central to infusion and immersion courses. In the infusion approach, CT improvement is explicitly stated as a goal, and students are aware that they are being trained to think critically in these classes. Conversely, in the immersion approach, students may not realise that they are receiving CT training. The mixed course combines the general approach with either the infusion or immersion approach.

The infusion approach to CT teaching has primarily been implemented in Kindergarten through 12th-grade (K-12) contexts (Ventura, Lai, & DiCerbo, 2017). This may be because it does not conflict with other educational objectives and is more readily accepted by schools as it does not add to an already crowded curriculum (Zohar & Tamir, 1993). Moreover, the infusion approach can be easily integrated into various disciplines (Bensley & Spero, 2014; Zulkpli, Abdullah, Kohar, & Ibrahim, 2017).

Some research suggests that the infusion approach can enhance students' CT skills (e.g., Bağ & Gürsoy, 2021; Zohar & Tamir, 1993; Zohar, Weinberger, & Tamir, 1994), though results should be interpreted cautiously. For example, Bağ and Gürsoy (2021) devised a CT-embedded English course and conducted a quasi-experimental study, showing beneficial effects on the CT skills of seventh-grade students in Turkey. However, the small sample size (31 per cell) limits the generalisability of their findings. Similarly, Lin (2014) conducted a case study on the infusion method in an English writing course at a Chinese public high school, indicating enhanced CT dispositions, skills, and language learning. However, methodological issues, such as the contamination problem and lack of appropriate comparators, weaken the evidence.

While most studies report positive effects of the infusion method (Ventura et al., 2017; Zulkpli et al., 2017), a few studies find no similar positive results (e.g., Toy & Ok, 2012). This inconsistency could be attributed to publication bias, where studies with promising results are more likely to be published (Song, Hooper, & Loke, 2013). For instance, Toy and Ok (2012) evaluated the infusion method in a vocational pre-service teacher education programme in Turkey, finding no particular benefit compared to no treatment.

Results examining the effects of infusion teaching of CT are inconsistent and inconclusive, likely due to variations in intervention duration, age range, and methodological differences. Nevertheless, previous reviews suggest that explicit methods, with strategies such as dialogue in class and the use of authentic problems or examples, are promising for CT skill development (Abrami et al., 2015).

In the context of China, the infusion method of CT appears to be the most relevant to the premise of this study as it is the most promising and can be easily embedded into the existing school curriculum without the need for additional add-on lessons. Thus, it is less likely to meet with resistance from schools in China, which are very exam- and textbook-oriented, as the infusion method can use the teaching resources and textbooks already used in schools. Furthermore, the most recent English curriculum standards for primary and secondary schools in China outline desired thinking outcomes for students, aligning with CT goals (MoE, 2022b). Integrating the infusion approach within the English curriculum is advantageous, given the mandatory nature of the subject.

The infusion method is widely used in K-12 education (Ventura et al., 2017), with secondary-aged (11-15 years old) students showing greater receptivity to CT instruction compared to postsecondary learners (Abrami et al., 2008). Hence, it is with these considerations that this study focuses on the infusion approach in English curriculum to teaching CT in secondary schools in China.

3. Research aims and questions

The primary objective of this study is to evaluate the impact of the infusion method of delivering CT on Chinese students' CT skills and academic attainment. Additionally, it aims to contribute to the debate on whether CT is a product of Western philosophy and therefore cannot be taught. Aligned with these objectives, the following research questions are raised:

1. Can critical thinking skills be taught to Chinese secondary students who are not traditionally exposed to critical thinking?
2. Does the infusion of critical thinking into the English curriculum improve Chinese secondary students' critical thinking skills?
3. Does the infusion of critical thinking into the English curriculum improve Chinese secondary students' academic performance?

4. Methods

4.1 The trial design

This study adopts a two-armed clustered randomised controlled trial (RCT), designed to answer causal research questions about the effectiveness of an intervention (Gorard, 2013). Due to the involvement of a comparison group and the randomisation process, RCTs are robust and reliable in investigating the causation between an intervention and outcome (Hariton & Locascio, 2018). The main trial's ethical application was approved by the School of Education Ethics Committee on 14th November 2022.

4.2 Participants and randomisation

The trial involved 2055 students in grade eight (aged 13-14 years old) from four secondary schools in the Sichuan province of China. The participants included 21 English language teachers, forming 21 clusters in total. These four secondary schools were situated in rural areas, somewhat disadvantaged compared to those in urban centres, with limited teaching facilities and resources. All schools were located in the same county. Since different English textbooks were used in various regions in China, students from the same area were exposed to the same curriculum content and tested by the same final examinations.

Randomisation occurred following a pre-test of CT skills and a student questionnaire on demographic information. Individual student randomisation was not feasible due to the intervention being delivered within schools. This was to respect the existing teaching structure (Gorard, 2021), where students were already assigned to different classes. Thus, randomisation was at the teacher level. Eleven teachers with their students (1,027) were randomised to deliver the intervention, while the remaining ten teachers with their students (1,028) formed the business-as-usual control group.

4.3 The intervention

The intervention was the infusion teaching of CT within the regular English curriculum. It spanned from March to May 2023 and was divided into two stages. Stage one consisted of seven lessons introducing the infusion method proposed by Elder and Paul (2020). Lessons 1-4 focused on teaching students intellectual standards (clarity and accuracy, relevance, depth and breadth, and logic). Lessons 5-7 centred on elements of reasoning, including evaluating information, identifying assumptions, and making inferences. Each lesson was delivered once a week, with a duration of 40-45 minutes per session. A sample lesson plan is provided in Appendix A.

Stage two involved six tasks designed to reinforce skills of analysis and evaluation, providing students with regular practice. These tasks integrated materials from the textbook with CT learning, offering students a different perspective on their usual English learning. As the CT tasks were linked to regular English teaching content and could be completed within approximately 20 minutes (less than a full lesson period), teachers were asked to deliver them based on their teaching pace. An example of a CT task is included in Appendix B.

The intervention was delivered in the usual classroom setting by the English teacher. To ensure fidelity, formal teacher training sessions and weekly informal follow-up training were conducted. Further details on teacher training are discussed in Fan (2024).

4.4 Outcomes

The primary outcome of this study was CT skills, assessed using question items from the Watson-Glaser Critical Thinking Appraisal, Cornell Critical Thinking Test Level X, and the Halpern Critical Thinking Test. While not all items from these standardised tests were used, they constituted a question bank from which items suitable for students of the target age group and relevant to their context were selected. Five CT sub-skills, including arguments, assumptions, deductions, inferences, and interpretation, were assessed. The pre-test of CT skills consisted of 15 multiple-choice questions, and students were required to complete them within 30 minutes to reduce test fatigue and boredom. To prevent familiarity with the test items, students were tested after the intervention with a post-CT test that measured the same CT sub-skills using different questions.

The secondary outcome, academic attainment, was measured by final examinations designed and administered by the local education department. Since Chinese, Maths, and English language are core subjects in secondary education in China, academic scores from these three subjects were collected as indicators of academic attainment.

Notably, all participating schools were from the same area, ensuring that students took the same tests and were blindly marked by teachers, ensuring consistency across schools.

4.5 Analysis

The impact of the intervention was assessed using Hedge's *g* effect size (ES), calculated as the mean difference in gain scores between the experimental and control groups divided by the overall standard deviation (SD). This metric provides an indication of the magnitude of differences between the two groups (Sullivan & Feinn, 2012). Gain scores were used to account for any baseline differences (if any) in CT skills and academic attainment between the experimental and control groups.

As is common in trial studies, there were missing cases in this study (Gorard, 2021). During the three-month intervention period, a total of 44 students (23 from the experimental group and 21 from the control group) were lost to follow-up. The primary reason for this attrition was that these students left the participating schools and could not be traced (e.g., transferring to new schools, suspension of schooling, or dropout). While it is unlikely that the attrition was caused by the intervention itself, as it was infused into the regular curriculum, it remains uncertain whether the experimental and control groups remained balanced as per the initial randomisation. Therefore, only completed cases were included in the analysis.

To assess whether the missing cases could potentially alter the findings, the number needed to disturb (NNTD) method was employed (Gorard & Gorard, 2016). This method evaluates whether the results would change if all missing cases were considered as the counterfactual ones (Gorard, 2021). The NNTD is calculated by multiplying the ES by the number of cases in the smallest group in the comparison, which in this study refers to the number of cases from the experimental group. If the NNTD is greater than the number of missing cases, it suggests that the results are robust and stable. Conversely, if the NNTD is smaller, caution should be exercised in interpreting the results.

4.6 Process evaluation

A process evaluation was conducted to assess implementation fidelity, understand the perspectives of teachers and students regarding the infusion teaching of CT, and identify any challenges or difficulties encountered by teachers during intervention delivery (Oakley, Strange, Bonell, Allen, & Stephenson, 2006; Gorard, See, & Siddiqui, 2017). The process evaluation included class observations and interviews with experimental teachers and students. Field notes and transcripts of interviews were coded to provide supplementary insights into the primary and secondary outcomes.

5. Findings

5.1 Impact evaluation on CT skills

Table 1 shows that the experimental group has a marginally better overall CT skill than the control group. Although the experimental group displays a higher score of the interpretation skill than the counterpart, these two groups are similar in terms of other skills including evaluating arguments, recognising assumptions, drawing deductions, and making inferences. This means that the experimental group is already slightly ahead of the control cohort. It is, therefore, necessary to include the gain score of each CT sub-set in the impact evaluation.

Table 1. Pre-test scores of critical thinking skills of the experimental and control groups (N = 2011)

	Experimental group (n = 1004)		Control group (n = 1007)		Overall (N = 2011)		ES
	Mean	SD	Mean	SD	Mean	SD	
Argument	1.48	0.89	1.55	0.86	1.51	0.88	-0.08
Assumption	2.15	0.79	2.08	0.86	2.12	0.83	0.08
Deduction	1.53	0.81	1.50	0.91	1.52	0.86	0.03
Inference	1.23	0.88	1.21	0.90	1.22	0.89	0.02
Interpretation	1.38	0.78	1.30	0.82	1.34	0.80	0.10
Overall CT skills	7.76	1.88	7.64	2.20	7.70	2.05	0.06

Experimental students appear to benefit from the infusion teaching of CT (see Table 2). Compared to students who did regular English learning in schools, they showed greater improvement in skills of identifying assumptions, making reasonable inferences and drawing logical deductions. On the other hand, changes in other CT skills including evaluating arguments and interpreting information remain stable between the two groups.

Table 2. Post-test scores of critical thinking skills between the experimental and control groups (N = 2011)

	Experimental group (n = 1004)		Control group (n = 1007)		Overall (N = 2011)		ES
	Mean	SD	Mean	SD	Mean	SD	
Argument	1.65	0.82	1.64	0.86	1.64	0.84	0.01
Assumption	2.38	0.75	2.21	0.84	2.29	0.80	0.21
Deduction	2.04	0.84	1.88	0.86	1.96	0.86	0.19
Inference	1.48	0.92	1.30	0.84	1.39	0.88	0.20
Interpretation	1.31	0.86	1.34	0.90	1.32	0.88	-0.03
Overall CT skills	8.85	2.16	8.36	2.33	8.60	2.26	0.22

To cater for any initial imbalances between groups, the gain scores analysis is also presented, and results are shown in Table 3. The intervention has a positive impact on the progress of the overall CT skills of these secondary students. Experimental students appeared to gain an improvement in skills of drawing inferences, making deductions, identifying assumptions and evaluating arguments.

However, they failed to make the same progress in the skill of interpretation as control students. It could be the case that the interpretation skill was not taught clearly in the CT infusion lessons. Other CT skills such as recognising assumptions and drawing reasonable inferences were delivered to students via the infusion CT lessons and reinforced in the CT tasks. However, there was no single lesson that focused on interpretation. Additionally, as these Chinese secondary students were still beginners in the English language, they did not have many opportunities to practice English in class. In most cases, they answered questions and discussed in groups in their native language (i.e., Chinese). It remains unknown whether this could influence the effectiveness of the intervention on students' progress in the interpretation skill.

Table 3. Gain scores of critical thinking skills between the experimental and control groups (N = 2011)

	Experimental group (n = 1004)		Control group (n = 1007)		Overall (N = 2011)		ES
	Mean	SD	Mean	SD	Mean	SD	
Argument	0.16	1.17	0.09	1.16	0.13	1.16	0.06
Assumption	0.24	1.00	0.12	1.11	0.18	1.06	0.11
Deduction	0.51	1.12	0.37	1.21	0.44	1.17	0.12
Inference	0.25	1.25	0.09	1.25	0.17	1.25	0.13
Interpretation	-0.07	1.13	0.04	1.18	-0.01	1.16	-0.09
Overall CT skills	1.09	2.59	0.71	2.90	0.91	2.76	0.14

5.2 Impact evaluation on academic attainment

Table 4 demonstrates that the experimental group were far behind their counterparts in all three subjects. Due to the imbalance in the prior academic attainment, it is reasonable to use the progress score, rather than the post-academic score, to evaluate the impact of the intervention on academic achievement. Besides, the English language is always the subject that students perform worse, compared to the other two subjects. This should be taken into consideration when CT is infused into the English language course. Perhaps a simpler English language should be used in the CT lessons to help students understand, apply and transfer the CT skills.

Table 4. Pre academic performance of the experimental and control groups

	Experimental group			Control group			Overall			ES
	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Chinese	1018	78.35	18.33	1018	86.71	14.63	2036	82.53	17.10	-0.49
Maths	1016	63.95	31.86	1017	76.14	30.89	2033	70.05	31.96	-0.38
English language	1017	56.91	26.02	1020	71.26	26.11	2037	64.09	27.03	-0.53

Overall academic scores	1015	199.46	68.71	1014	234.43	64.81	2029	216.94	69.03	-0.51

According to the post-academic scores, it seems that experimental students did not benefit from the intervention in terms of academic achievement (see Table 5). However, this was tentative as the academic performance of these two groups was not balanced at the outset.

Table 5. Post academic performance of experimental and control groups

	Experimental group			Control group			Overall			ES
	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Chinese	1000	78.42	19.19	1006	86.95	14.74	2006	82.74	17.63	-0.48
Maths	999	60.07	29.36	1005	72.59	28.83	2004	66.33	29.79	-0.42
English language	1000	59.83	27.48	1007	74.76	28.04	2007	67.30	28.79	-0.52
Overall academic scores	992	198.43	67.74	1004	234.30	64.70	1996	216.37	68.62	-0.52

Table 6 shows that the control students increased their average Chinese score while the experimental students showed a decrease. Both groups failed to make progress in Maths, and the experimental group obtained a lower mean score than the business-as-usual cohort. On the other hand, both groups demonstrated an increase in the English language subject. However, the progress of experimental students was not as good as that of control students.

Overall, the intervention’s impact on the academic performance is not very promising, but it should be treated with caution because the experimental group had already been behind the control group at the outset. Alternatively, students’ academic attainment might be improved if the intervention was conducted longer enough. It is also possible that CT skills were not tested in the final examinations.

Table 6. Gain academic performance of experimental and control groups

	Experimental group			Control group			Overall			ES
	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Chinese	996	-0.31	10.99	1000	0.24	9.68	1996	-0.03	10.36	-0.05
Maths	994	-4.55	13.59	998	-3.76	13.75	1992	-4.15	13.67	-0.06
English language	996	2.43	10.57	1003	3.31	9.49	1999	2.87	10.05	-0.09
Overall academic scores	987	-2.36	21.96	994	-0.26	21.25	1981	-1.30	21.63	-0.10

5.3 Sensitivity

To evaluate the degree of security of the results, a sensitivity analysis was employed (Thabane et al., 2013). As mentioned earlier, the NNTD is calculated as the 'effect' size multiplied by the number of cases in the smallest group in the comparison. As the smaller cell in this main trial is the experimental group that includes 1004 students and the effect size of CT skills' gain score is +0.14, the NNTD is 141 (1004*0.14). This means that we need all 141 cases to have negative results to change the effect. If the number of missing cases is larger than 141, the findings are insecure because the missing cases would alter the findings if they were considered. However, the real missing cases of CT skills testing, as described in the achieved sample, is 44, which takes up only 31% of NNTD of 141. Therefore, the results of CT effect size are secured.

Although the missing cases are unlikely to alter the final findings, an investigation of their pre-test score of CT skills gives us a better understanding of the effect of the intervention. Table 7 shows that the missing post-test scores from the experimental group have higher pre-test scores. While the difference in the scores of arguments and assumptions between experimental and control groups is small, experimental students who do not have complete post-test scores perform better in other CT facets including deduction, inference and interpretation. Therefore, despite the balanced number of attritions in each group, students who dropped out of the experimental group are more likely to be higher critical thinkers. This may pull down the overall impact.

Table 7. A comparison of the pre-test CT scores of the missing cases (N = 44)

	Experimental group (n = 23)		Control group (n = 21)		Overall (N = 44)		ES
	Mean	SD	Mean	SD	Mean	SD	
Argument	1.35	0.83	1.43	0.93	1.39	0.87	-0.09
Assumption	1.91	0.85	1.86	0.85	1.89	0.84	0.06
Deduction	1.57	0.84	1.29	0.85	1.43	0.85	0.33
Inference	1.35	0.88	0.95	0.86	1.16	0.89	0.45
Interpretation	1.35	0.71	1.19	0.87	1.27	0.79	0.20
Overall CT skills	7.52	1.68	6.71	2.19	7.14	1.96	0.41

Table 8 provides the NNTD of each subject. The effect size of overall academic scores was -0.10, and the smaller cell of this main trial was the experimental cohort that included 987 students. Hence, the NNTD is 99 (987*0.10). That is to say, if the final results were to be altered, we need all 99 cases to be counterfactual. Since the real number of missing cases in the overall academic scores was 65, which is smaller than 99, it could be safely concluded that the results were robust and trustworthy.

However, if the specific academic subjects were considered, the intervention impact on the Chinese and Maths skills of students should be read with caution. The result of the English language tended to be more robust because of the smaller number of missing cases as opposed to NNTD. In sum, the attrition did not unduly influence the substantive results in English and overall academic scores, but the results of the intervention impact on the gain scores of Chinese and Maths should be treated with caution.

Table 8. Results of the sensitivity analysis of academic attainment

	Smaller cell	Number of missing cases	Effect size	NNTD
Chinese	996	50	-0.05	50
Maths	994	54	-0.06	60
English language	996	47	-0.09	90

Overall academic scores	987	65	-0.10	99

5.4 Process evaluation

Class observations: experimental and control

The observation of the experimental class suggests that the intervention was fairly implemented. In most cases, teachers taught the infusion CT course as planned. Indeed, in the first two lessons, teachers were not familiar with the intervention and did not know what the CT course looked like. Most teachers said that they were not used to teaching CT because they did not give students correct answers. Things were much better when the intervention was implemented longer. Teachers understood that they were facilitators in class, and should give students time to think, discuss and share ideas. As one teacher said, "*Let students lead the class, and they will surprise you*". They tried not to directly give answers and explain for students. For example, teachers could ask some thought-provoking questions, and students demonstrated a good capability of justifying their answers.

Students' engagement was observed as high, which is consistent with their responses in the interviews. Their dialogues and answers in classes suggested that the intervention had a positive impact on their CT skills. For example, in the third infusion lesson, students were divided into two groups to consider the pros and cons of having shorter school days. When one group was sharing their answers, the other was judging whether the point was relevant, non-repetitive, and convincing. If not, they would consciously argue against it, as the following conversation indicates.

Student 1: If we have shorter school days, we will fall behind in learning.
Student 2: I want to argue against Student 1's point. The learning time is different from the learning outcome. If we have more learning time, our pressure would be more.
Student 1: I want to argue against Student 2's view. We could be more motivated if we have pressure.

The usual English classes delivered by the control teachers were also observed. These classes put much emphasis on the English language, and CT was not a focus. This is common because the main task for English teachers was to finish teaching the textbook and prepare students for the tests. The teaching approach was teacher centred. According to them, it would be a waste of time to cultivate CT because it was not tested.

<u>*Views of experimental students*</u>

Most students have never been exposed to CT before. When they heard that they were going to learn the infusion CT course, they thought it was *"like giving you puzzles and then asking you to complete it"*, and teachers would *"teach complicated and abstract ideas that students do not understand"*. However, after learning the CT lessons and tasks, they realised that it was different from what they had imagined. They found the intervention was interesting. One reason for this was that the adopted examples and exercises were close to their daily life so they could easily relate the CT content to their previous experiences and knowledge. Another reason was the change in teaching style. Some students commented *"The infusion CT course is somewhat different from the traditional way of teaching. Traditional education is teacher-centred, and this is one-sided input for students. However, this CT course is a two-way communication. It is a mutual cooperation that searches for the truth."*

The infusion CT lessons were also perceived to be relaxing. There was less workload assigned to students. They did not have to memorise new words and grammar knowledge, and there was no homework afterwards. Most of the questions were open-ended, and students were allowed to provide their own answers. They were not worried about giving wrong answers and being criticised.

Most students were convinced that their CT skills had been improved. They argued that they could consider the relevance of conversations, judge some information, reflect common assumptions, and distinguish correlation and causation. As one student reflected,

> *"I used to accept all arguments without thinking about them carefully. It did not matter for me to think if these arguments were right. It seemed like it had no effect on me anyway…. People said that girls were born to be good at liberal arts, while boys were bad at them, and good at science. When you play on the*

> *phone, you would have poor eyesight, and then your academic grades must decrease. It seems that the result is of 100% certainty. Previously, I did not care about it, and it seemed not to affect me. But this time, I think it is different, and I think they are wrong. These factors and results were just correlated, why do people assume there was a causal relationship (among these arguments)?"*

However, they held different opinions on the intervention's impact on their academic learning. Some believed the infusion CT course inspired them to look for alternative answers, write compositions, analyse textbook articles, expand their vocabulary, and enhance their enthusiasm for English learning. Nonetheless, some thought it was not very helpful in improving their academic learning. This may be because they had no idea how to use the CT lesson content in their regular learning. Another reason is that they assumed the CT course was not relevant to their academic study. Therefore, they showed little interest and were less engaged in class.

Experimental teachers' opinions

The experimental teachers did not have any experience teaching CT to students. They thought the infusion CT course would be abstract and complicated at first. Some of them only heard of the term "critical thinking" but had no idea of what it was. However, after the teacher training and the progress of CT lessons, they felt much easier to teach and more familiar with the CT teaching.

All teachers liked the infusion CT course and believed it was interesting and interactive. They noticed the difference in teaching approaches between the CT course and the usual English class. As some teachers commented,

> *"The usual English class was more serious. Sometimes students felt (the English course content) was too abstract to understand, but we teachers had no way. We had to finish our teaching task, so I had to cram knowledge into them regardless of whether they understood it. This is spoon-feeding education."*

All teachers agreed that the intervention had improved their students' thinking skills. One teacher gave an example of students judging the trustworthiness of a video in a regular English class.

Teacher: I think (the infusion CT course) had an influence (on students' thinking). A good example was that I played a video to students yesterday. My intention was to ask them to watch the video and learn about its content. But all of them said it was fake. They were judging the credibility of the video. Even if I told them that it was true, they still said it was fake. I think it really affected their thinking. Now I hand out course materials to them, and they will think about whether the material is right rather than (directly) analysing it.
I: If they directly gave the conclusion, will you ask them 'How do you come to this conclusion'?
Teacher: I did not ask them. They consciously gave reasons why (they thought) it was fake. They said, 'Look at the video. The person looks fake, and the building suddenly disappears. It is made of three-dimensional animation'.

However, teachers were hesitant to say that there was a positive impact on the students' English learning. While some teachers mentioned it could help students understand articles better and use some sentences in writing compositions, most thought the influence, if any, was minimal. One reason was that the infusion CT course was not for English tests. According to one teacher, *"A large part of our (English) test is reading questions. If students use what has been learned from the thinking course, it is difficult for them to choose the right answers."* Some teachers also thought it would take a long time to make the influence on students' English learning obvious. This indicates that the intervention might be implemented longer.

It is interesting to note that all teachers anticipated that the infusion CT course might have a better influence if it is taught in more advanced areas. They maintained that their schools were in the countryside, and English course was not taught as seriously as those in Chinese and Maths subjects in primary schools. Students did not have the opportunity to learn English at an early age, and their students were not good at English. Besides, students from urban areas were perceived to be more likely to access more educational resources and thus have more background knowledge.

6. Discussions and conclusion

The trial findings indicate that the infusion of CT has been well received by both teachers and students in China. Concerns regarding the link between CT skills and academic assessments were noted, yet there was no resistance to the delivery or reception of CT lessons. Feedback from interviews with teachers and students was positive, highlighting high levels of engagement and enthusiasm for the infusion CT lessons. The successful implementation of CT content within English classes, facilitated by a range of teaching resources demonstrates the feasibility and popularity of the infusion approach. Additionally, suggestions to formally integrate CT into the English curriculum and expand its implementation to other educational levels further endorse the appropriateness and promise of the intervention in China. These findings challenge Atkinson's (1997) assertion that CT cannot be taught to individuals from non-Western cultures and align with Lin's (2014) study, where CT was successfully infused into the English curriculum of Chinese high schools.

The small positive effect size of +0.14 suggests that the infusion CT approach has the potential to enhance CT skills among Chinese secondary school students. This finding is consistent with previous studies demonstrating the effectiveness of the infusion method of CT teaching (e.g., Bağ & Gürsoy, 2021; Zohar & Tamir, 1993; Zohar et al., 1994). This finding also provides evidence that CT skills could be improved when the intervention is conducted for three months (Niu, Behar-Horenstein, & Garvan, 2013). Furthermore, the trial provides evidence of the positive impact of the infusion method on skills such as argumentation, assumption identification, deduction, and inference. The effectiveness of the intervention may be attributed to the well implementation of the infusion teaching (Harn, Parisi, & Stoolmiller, 2013; O'Donnell, 2008).

However, the infusion course did not lead to an increase in students' interpretation skills. This could be attributed to the possibility that interpretation skill development requires a longer period to be evident. Alternatively, it may be due to unclear instruction on a specific interpretation skill question, such as understanding the distinction between correlation and causation, as observed during class observations. This suggests a need for clearer instruction and reinforcement of interpretation skills in future interventions.

It appears that students who participated in the infusion CT course did not show better progress in Chinese, Maths, and English scores. This contrasts with Lin's (2014) study, where the progress in English writing was observed among 89 Chinese high school students following CT infusion teaching. Similarly, Hu et al. (2011) found that exposure to thinking teaching led to improved academic attainment, including Chinese and Maths, among 116 Chinese primary school students. However, caution is warranted when interpreting the trial's results on academic attainment, as dropout cases and initial imbalances in academic performance between experimental and control groups may have influenced the evaluation. The potential bias of class segregation, where students with similar academic performance are grouped together, could also have affected the outcomes, given that randomisation occurred at the teacher level.

While the infusion CT approach demonstrated a small positive impact on students' CT skills, this effect did not extend to academic attainment. This suggests that improving academic performance may require a longer intervention period or a more intensive intervention dose than enhancing CT skills alone. For instance, as a meta-analysis shows, the association between CT and academic achievement is stronger when assessed one year later (Fong, Kim, Davis, Hoang, & Kim, 2017). Another plausible explanation could be that CT skills are rarely tested in the academic assessments in China. Teachers from the trial were concerned that they had less time to finish the English textbook teaching in class. Their concerns indicated that CT skills, or at least the infusion CT content, were not included in the academic assessments. This seems to be aligned with the rote learning of Chinese students and knowledge-oriented assessments in China. Even if some university instructors taught the CT course, the summative test is to check if some rules of thinking are memorised (Dong, 2015). Therefore, a reform of academic assessments is needed if CT skills are to be cultivated.

In conclusion, while the results of the randomised controlled trial support the feasibility and promise of infusing CT into the English curriculum in Chinese secondary schools, further efforts are necessary. This aligns with findings from previous studies (Bağ &

Gürsoy, 2021; Lin, 2014; Zohar & Tamir, 1993). While advancements in CT skills have been outlined in the English curriculum standards for secondary schools (MoE, 2022b), this represents only the initial stage of implementation. Continued efforts are required:

- Provide practical teacher training on CT teaching as a national initiative: Instead of giving lectures on theories of CT, the teacher training should be more practical and equipped with hands-on teaching resources.
- Allocate more teaching time for teachers: While the infusion approach can be appropriate for school timetabling, it still takes time to complete CT-related activities, which reduces the time for obtaining curricular knowledge (Solon, 2007).
- Include CT skills in academic assessments: This is to cater for the issue of teaching to the test in China. It could raise teachers, students and even parents' awareness of the importance of CT. The testing of CT skills should go beyond the memorisation of CT theories. More serious and conscious efforts are needed to design a CT-related academic assessment in China.
- Reform textbooks to accommodate CT infusion teaching: The amount and difficulty level of the content should be updated.
- Invest teachers with more autonomy and flexibility: CT teaching is rare in primary, secondary and high schools in China as teachers have to complete the regulated teaching content, and prepare their students for the high-stakes assessment each term.
- Reduce teachers' workload: Chinese teachers are carrying out many irrelevant teaching issues such as administrative activities (Liu & Onwuegbuzie, 2012; Yan, 2015). The reduction of their workload allows them to focus on the teaching of subjects and CT.
- Have a small class size: This makes teachers devote enough time to each student and students have more time to think about issues, discuss with peers and share ideas. It is also easier for teachers to create a comfortable class atmosphere and let students express their ideas freely.

Limitations and future research

The limitations of this study highlight several avenues for future research to address. Firstly, future studies should consider randomising students at the individual level rather than the cluster level to deal with issues related to class segregation. This approach would enhance the balance between groups and provide more robust evidence regarding the intervention's impact.

Secondly, despite the large sample size, efforts should be made to address the dropout of students to ensure a more comprehensive evaluation of the intervention's impact on academic performance. Strategies such as implementing measures to reduce dropout rates or conducting follow-up assessments with students who leave the study could be helpful.

Thirdly, future research should explore the effectiveness of CT infusion teaching in schools with different demographic characteristics, including those in urban areas with students who have earlier exposure to English language learning. Students from this area did not regularly learn English until going to secondary schools. Their English language level may be at a disadvantage, compared to students from urban areas who start learning English at an early age (e.g., in primary schools or even kindergarten). As the extent of using English to instruct in the infusion courses depends on students' English language proficiency, it is unknown whether students' English skills would be improved if more English were used in the infusion lessons. In addition, these participating students seemed to have lower socioeconomic status than those from urban schools. Although the inclusion of rural secondary schools in China could enrich the current CT education in China that is predominantly focused on urban schools, the generalisation of the findings from this study should be made cautious.

Finally, expanding the evaluation beyond the English curriculum at secondary schools to include other academic disciplines and educational levels is essential. This could offer a more comprehensive understanding of the impact on students' CT skills and academic attainment and help establish the broader applicability of the intervention.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. R*eview of educational research, 85*(2), 275-314.

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102-1134.

Atkinson, D. (1997). A Critical Approach to Critical Thinking in TESOL. *TESOL Quarterly, 31*(1), 71-94. https://doi.org/10.2307/3587975

Bağ, H. K., & Gürsoy, E. (2021). The effect of critical thinking embedded English course design to the improvement of critical thinking skills of secondary school learners. *Thinking Skills and Creativity, 41*, 1-13.

Bailin, S. (2002). Critical thinking and science education. *Science & Education, 11*(4), 361–375.

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

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

Bensley, D. A., & Spero, R. A. (2014). Improving critical thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity, 12*, 55–68.

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

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

Cheng, Y. C., Huang, L. C., Yang, C. H., & Chang, H. C. (2020). Experiential learning program to strengthen self-reflection and critical thinking in freshmen nursing students during COVID-19: A quasi-experimental study. *International journal of environmental research and public health, 17*(15), 5442.

Cui, R., & Teo, P. (2023). Thinking through talk: Using dialogue to develop students' critical thinking. *Teaching and Teacher Education*, 125, 104068.

Dello-Iacovo, B. (2009). Curriculum reform and 'quality education' in China: An overview. *International journal of educational development, 29*(3), 241-249.

Dhakal, K. R., Watson Todd, R., & Jaturapitakkul, N. (2023). Unpacking the nature of critical thinking for educational purposes. *Educational Research and Evaluation*, *28*(4-6), 130-151.

Dong, Y. (2015). Critical Thinking Education with Chinese Characteristics. In Davies, M., Barnett, R. (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 351-368). Palgrave Macmillan, New York. https://doi.org/10.1057/9781137378057_22

Dong, Y. (2017). Teaching and assessing critical thinking in second language writing: An infusion approach. *Chinese Journal of Applied Linguistics*, *40*(4), 431–451. https://doi.org/10.1515/cjal-2017-0025

Elder, L., & Paul, R. (2020). *Critical thinking: Tools for taking charge of your learning and your life.* Foundation for Critical Thinking.

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* (pp.9-26). New York: W.H. Freeman and Company.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational researcher*, *18*(3), 4-10.

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

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.

Fan, K. (2024). Are Chinese students deficient in critical thinking? In S. Gorard & N. Siddiqui (Eds.), *An International Approach to Developing Early Career Researchers: A Pipeline to Robust Education Research.* London: Routledge.

Fan, K., & See, B. H. (2022). How do Chinese students' critical thinking compare with other students?: a structured review of the existing evidence. *Thinking Skills and Creativity, 46,* https://doi.org/10.1016/j.tsc.2022.101145

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

Fong, C. J., Kim, Y., Davis, C. W., Hoang, T., & Kim, Y. W. (2017). A meta-analysis on critical thinking and community college student achievement. *Thinking Skills and Creativity*, *26*, 71-83.

Fung, D. (2017). The pedagogical impacts on students' development of critical thinking dispositions: Experience from Hong Kong secondary schools. *Thinking Skills and Creativity,* 26, 128-139.

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

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

Gorard, S., & Gorard, J. (2016). What to do instead of significance testing? Calculating the 'number of counterfactual cases needed to disturb a finding'. *International Journal of Social Research Methodology, 19*(4), 481-490. https://doi.org/10.1080/13645579.2015.1091235

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

Halpern, D. F. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New directions for teaching and learning*, *1999*(80), 69-74.

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: an international journal of obstetrics and gynaecology*, *125*(13), 1716.

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools?. *Exceptional Children*, *79*(2), 181-193.

Horn, S., & Veermans, K. (2019). Critical thinking efficacy and transfer skills defend against 'fake news' at an international school in Finland. *Journal of Research in International Education, 18*(1), 23-41.

Hu, W., Adey, P., Jia, X., Liu, J., Zhang, L., Li, J., & Dong, X. (2011). Effects of a 'Learn to Think' intervention programme on primary school students. *British journal of educational psychology, 81*(4), 531-557.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice, 41*(4), 212-218.

Ku, K. Y., Ho, I. T., Hau, K. T., & Lai, E. C. (2014). Integrating direct and inquiry-based instruction in the teaching of critical thinking: an intervention study. *Instructional Science*, 42, 251-269.

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports, 6*, 1-49.

Li, X. (2017). Exploration and Reflection in Dialogue and Development—The exploration and practice of cultivating pupils' critical thinking in the Affiliated Primary School of Huazhong University of Science and Technology. *New Curriculum Research,* (6), 31-33.

Lin, Y. (2014). *Infusion of critical thinking into L2 classes: a case study in a Chinese high school.* (Doctoral dissertation, Newcastle University).

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

Liu, S., & Onwuegbuzie, A. J. (2012). Chinese teachers' work stress and their turnover intention. *International journal of educational research*, *53*, 160-170.

Marin, L., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, 6, 1-13.

Ministry of Education. (2022a). *Compulsory Education Curriculum Plan (2022 edition).* Retrieved from http://www.moe.gov.cn/srcsite/A26/s8001/202204/W020220420582343217634.pdf Accessed on 3 Mar 2024.

Ministry of Education. (2022b). *English curriculum standards for compulsory education.* Retrieved from http://www.moe.gov.cn/srcsite/A26/s8001/202204/W020220420582349487953.pdf Accessed on 18 Jun 2022.

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

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational research review, 9*, 114-128.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of educational research*, *78*(1), 33-84.

Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Process evaluation in randomised controlled trials of complex interventions. *Bmj*, *332*(7538), 413-416.

OECD. (2011). *Shanghai and Hong Kong: Two Distinct Examples of Education Reform in China. in Lessons from PISA for the United States,* OECD Publishing, Paris, https://doi.org/10.1787/9789264096660-5-en.

OECD. (2018). *Future of education and skills: Skills for 2020.* Paris: OECD. Retrieved from https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf Accessed on 18 Jun 2022.

Paul, R., & Elder, L. (2010). *The Miniature Guide to Critical Thinking Concepts and Tools.* Dillon Beach: Foundation for Critical Thinking Press.

People's Republic of China. (2010). *Outline of China's national plan for medium and long-term education reform and development (2010-2020).* Beijing: People's Publishing House.

Solon, T. (2007). Generic critical thinking infusion and course content learning in introductory psychology. *Journal of Instructional Psychology*, *34*(2), 95-109.

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.

Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement*. National Institute of Education. http://eric.ed.gov/PDFS/ED272882.pdf

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. J*ournal of graduate medical education, 4*(3), 279-282.

Tan, C. (2020). Conceptions and practices of critical thinking in Chinese schools: An example from Shanghai. *Educational Studies, 56*(4), 331-346.

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., ... & Goldsmith, C. H. (2013). A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC medical research methodology, 13*(1), 1-12.

Toomey, E., Hardeman, W., Hankonen, N., Byrne, M., McSharry, J., Matvienko-Sikar, K., & Lorencatto, F. (2020). Focusing on fidelity: narrative review and recommendations for improving intervention fidelity within trials of health behaviour change interventions. *Health Psychology and Behavioral Medicine, 8*(1), 132-151.

Toy, B. Y., & Ok, A. (2012). Incorporating critical thinking in the pedagogical content of a teacher education programme: does it make a difference?. *European Journal of Teacher Education, 35*(1), 39-56.

Tsipursky, G., & Ward, T. (2020). *Pro truth: A practical plan for putting truth back into politics.* John Hunt Publishing.

Ventura, M., Lai, E., & DiCerbo, K. (2017). *Skills for Today: What We Know about Teaching and Assessing Critical Thinking.* London: Pearson.

Wang, H. H., Chen, H. T., Lin, H. S., Huang, Y. N., & Hong, Z. R. (2017). Longitudinal study of a cooperation-driven, socio-scientific issue intervention on promoting students' critical thinking and self-regulation in learning science. *International Journal of Science Education, 39*(15), 2002-2026.

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, 8-19.

Yan, C. (2012). 'We can only change in a small way': A study of secondary English teachers' implementation of curriculum reform in China. *Journal of Educational Change*, 13, 431-447.

Yan, C. (2015). 'We can't change much unless the exams change': Teachers' dilemmas in the curriculum reform in China. *Improving Schools*, *18*(1), 5-19.

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

Zhai, J.Y. (2015). Critical thinking curriculum, why claim that Chinese students really need it?. *Critical and Creative Thinking Education Newsletter, 24*, 24-27.

Zhang, C., Fan, H., Xia, J., Guo, H., Jiang, X., & Yan, Y. (2017). The effects of reflective training on the disposition of critical thinking for nursing students in China: A controlled trial. *Asian Nursing Research, 11*(3), 194-200.

Zhao, W. (2020). Epistemological flashpoint in China's classroom reform: (How) can a 'Confucian do-after-me pedagogy' cultivate critical thinking?. *Journal of Curriculum Studies, 52*(1), 101-117.

Zohar, A., & Tamir, P. (1993). Incorporating critical thinking into a regular high school biology curriculum. *School Science and Mathematics, 93*(3), 136-140.

Zohar, A., Weinberger, Y., & Tamir, P. (1994). The effect of the biology critical thinking project on the development of critical thinking. *Journal of Research in Science Teaching, 31*(2), 183-196.

Zou, M., & Lee, I. (2023). Learning to teach critical thinking: testimonies of three EFL teachers in China. *Asia pacific journal of education, 43*(3), 867-881.

Zulkpli, Z., Abdullah, A. H., Kohar, U. H. A., & Ibrahim, N. H. (2017). A review research on infusion approach in teaching thinking: advantages and impacts. *Man India.*, *97*(12), 289-298.

Author statement

Keji Fan: The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.