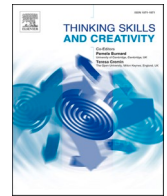


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Thinking Skills and Creativity

journal homepage: www.elsevier.com/locate/tsc

Effects of problem-based learning instructional intervention on critical thinking in higher education: A meta-analysis

Yong Liu ^{a,*}, Attila Pásztor ^{b,c,d}

^a Doctoral School of Education, University of Szeged, Hungary

^b Institute of Education, University of Szeged, Hungary

^c MTA-SZTE Research Group on the Development of Competencies, Hungary

^d MTA-SZTE Digital Learning Technologies Research Group, Hungary

ARTICLE INFO

Keywords:

Problem-based learning
Critical thinking
Intervention
Higher education
Meta-analysis

ABSTRACT

In recent years, the debate has continued among researchers and instructors regarding the influence of Problem-Based Learning (PBL) on the effectiveness of instructional intervention for Critical Thinking (CT) in higher education. This study, conducting a meta-analysis by synthesizing 50 relevant empirical studies from 2000 to 2021 with 5,210 participants and 58 effect sizes, aims to present potential factors (i.e., sample size, sample type, instruction type, gender, maturity, instrument, nationality, discipline, treatment duration, and group size) that may influence the effectiveness of the cultivation of CT skills and disposition on the basis of PBL. No evident publication bias was found (Egger's bias = 1.21, $p > 0.05$). From the general perspective, the results demonstrate the high level of influence of PBL (Standardized Mean Difference [SMD] = 0.640, $p < 0.001$) on CT with heterogeneity ($I^2 = 82.9\%$) due to the adopted instruments, mixed methods, and target outcomes, and no difference was observed between influence on CT skills and disposition. Students' maturity, nationality, sample type, instruction type, and group size are influencing factors of overall CT. The effects of intervention for seniors, western students, randomized samples, online instruction, and groups with less than six members are better, whereas short-term intervention is ineffective. For CT skills, the treatments for juniors and groups with less than six members are ineffective, and sample type and instruction type are not influencing factors. However, the effect sizes of big sample sizes, seniors, other kinds of instruments, western students, Sciences as a discipline, more than ten members in a group, and long-term intervention are stronger. For CT disposition, sample type, instruction type, discipline, and intervention duration are influencing factors, in which randomized samples, online instruction, students in Medicine, and medium-term intervention exerted a stronger effect than the other factors. In conclusion, although PBL is overall effective for promoting the acquisition of CT (skills and disposition), additional studies are also required to explore the effectiveness and influencing factors in other contexts, such as various learning or teaching strategies, environments, and scaffoldings, and scenario-problem-based tasks instead of only curriculum-based ones. These factors should also be considered to promote CT skills and disposition among undergraduates.

* Corresponding author at: Doctoral School of Education, University of Szeged, Petőfi S. sgt. 32-34, Szeged, 6722 Hungary.
E-mail address: yong.liu@edu.u-szeged.hu (Y. Liu).

<https://doi.org/10.1016/j.tsc.2022.101069>

Received 7 December 2021; Received in revised form 19 May 2022; Accepted 10 June 2022

Available online 11 June 2022

1871-1871/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Critical thinking (CT), as certain higher-order thinking, has been regarded as a planned achievement of education in 2050 (International Commission on the Futures of Education Commission, 2021), which, when taught effectively, will promote logical problem-solving (Dwyer et al., 2011) and contribute to the educational improvement, especially in higher education, and the job market. However, teaching CT to undergraduates is a major academic challenge (Kuhn, 1991; Willingham, 2010) because of the difficulties of embedding CT into an existed well-organized curriculum, infusing some valid tasks, and utilizing an effective teaching strategy (Dwyer et al., 2011). Problem-Based Learning (PBL), a student-centred approach, emphasizing learning by solving problems, has been suggested and used at the university level for developing CT in undergraduates. Most studies demonstrated the overall positive effects while some were negative. Thus, an elaborated literature review and meta-analysis are necessary to explore the inconsistency, overview the effectiveness, and detail the influencing factors of the effectiveness of interventions.

1.1. Definition of CT

CT has been considered one of the skills required in the 21st century (Trilling & Fadel, 2009), the facet of global competencies, and a core skill for 2030 (OECD, 2018). However, providing a standard definition of CT is a challenging endeavour due to the complexity of CT as a psychological construct. However, various definitions outline similar broad dimensions. For example, CT, as reasoned, reflective thinking focused on deciding what to believe or do (Ennis, 2011), pertains to the capacity to evaluate statements (Lawson, 1999). In addition, it is deemed the purpose of instruction, in which students apply cognitive skills, such as forming hypotheses, designing, performing, and analysing a series of investigations (Dell'Olio & Donk, 2007; Gomez, 2002; Wiles & Bondi, 1989). Despite its complexity, most scholars (e.g., Elder & Paul, 2020; Ennis, 2011; Facione, 1990a; Halpern, 1998) agree that CT is dependent on skills and disposition.

Component skills that involve CT have been frequently discussed issues in the academe. The American Philosophical Association outlined six primary CT skills (Facione, 1990a): interpretation, analysis, evaluation, inference, explanation, and self-regulation. These skills laid the foundation for later versions, such as the dominantly proposed skills categories: (a) focusing on issues; analysing arguments; posing questions and answering; clarifying and challenging questions; judging the credibility of sources; and observing and judging assumptions (Ennis, 2011); (b) verifying hypotheses; providing verbal reasoning; identifying uncertainty and making decisions; and offering solutions to problems and creativity (Halpern, 1998); (c) utilizing creative thinking and CT; practicing decision-making; and solving everyday and mathematical problems (Perkins et al., 1993); and (d) conducting analysis, inference, induction, and evaluation (Adler, 2000). These skills are identified as crucial elements of CT.

Disposition, which is related to motivation and attitude toward acquiring CT skills, is regarded as a process that activates skills (Ennis, 2011; Norris, 2003; Perkins et al., 1993) and as attitudes or consolidated intellectual habits (Paul & Binker, 1990; Salomon, 1994). The consensus of the contents of disposition toward CT regards it as characterological or intellectual attributes, such as truth-seeking, open-mindedness, analyticity, systematicity, self-confidence, inquisitiveness, and maturity of judgment (Facione et al., 2000a, 2000b). Nevertheless, empirical studies on the disposition toward CT or the motivational aspects of this manner of cognition are scarce.

1.2. Assessment of CT

Several CT skills and disposition assessment tools have been formulated to evaluate and promote CT for undergraduates based on the dimensions above.

1.2.1. Assessment of CT skills

Collegiate Learning Assessment Plus (CLA+), developed by the Council for Aid to Education in 2013, focuses on analysis and problem-solving, writing effectiveness, writing mechanics, scientific and quantitative reasoning, critical reading and evaluation, and critique of an argument. The HEIghten Outcomes Assessment Suite, created by Educational Testing Service in 2015, is embedded in analytic and synthetic skills. The Halpern Critical Thinking Assessment (HCTA) summarized five scenarios of CT, namely, verbal reasoning, argument analysis, thinking as hypothesis testing, using likelihood and uncertainty, and decision-making and problem-solving skills (Butler et al., 2012). California Critical Thinking Skills Test (CCTST), invented by Facione (1990a), comprises five subscales: analysis, evaluation, inference, deductive reasoning, inductive reasoning, and total CT skills. Another scale is the Watson–Glaser Critical Thinking Appraisal (WG-CTA; Watson, 1980), based on five subscales: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments (Bernard et al., 2008). Moreover, many studies adopted other or self-developed instruments, such as the Cornell Critical Thinking Test (CCTT; Ennis et al., 1964) and Thinking Skills Assessment (TSA, developed by world-leading universities), to explore the effectiveness of teaching CT skills.

1.2.2. Assessment of CT disposition

The first-invented one, California Critical Thinking Disposition Inventory (CCTDI; Facione & Facione, 1992), based on the Delphi Report (Facione, 1990b), measures seven attributes that influence the capacity to learn and to apply CT skills effectively, namely, truth-seeking, open-mindedness, the anticipation of possible consequences, proceeding systematically, confidence in one's power of reasoning, inquisitive to learning, and mature judgment. Furthermore, Yoon (2004) developed Yoon's Critical Thinking Disposition scale for Korean nursing students and presented the dimensions relatively similar to CCTDI: objectivity, prudence, systematicity,

intellectual curiosity, intellectual fairness, healthy skepticism, and self-confidence. [Sosu \(2013\)](#) developed and validated the new CT Disposition Scale with two factors, namely, critical openness and reflective skepticism. [Quinn et al. \(2020\)](#) developed and validated the Student–Educator Negotiated CT Dispositions Scale, which aims to fill the gap where instruments are only involved in expert and academic discussions without considering students’ perceptions. The scale covers six dimensions: reflection, attentiveness, open-mindedness, organization, perseverance, and intrinsic goal motivation.

Despite the existence of fruitful instruments, their effectiveness as used in different studies displayed varied results, which leads to the necessity to explore the effect size of each tool.

1.3. CT instructional intervention for higher education

Education on CT remains controversial and confusing for many instructors ([Bensley & Murtagh, 2012](#)); however, consensus exists among researchers that CT can be taught and learned ([Halpern, 2001](#)). Students may improve their CT ability with the teachers’ use of appropriate pedagogies and curriculum materials ([Godzella & Masten, 1998](#); [Halpern, 2001](#); [McMillan, 1987](#)), active learning strategies ([Kim, 2009](#)), and interactions among students and between students and instructors ([Cooper, 1995](#); [Howe & Warren, 1989](#)). Thus, although most experts cannot demonstrate the effectiveness of specific approaches, instructional interventions are essential.

For instructional interventions for CT, previous investigations demonstrate two paths, namely, (a) a programmatic path through the entire curriculum of a degree program and (b) an instructional path, which pertains to specific instructional approaches ([Behar-Horenstein & Niu, 2011](#)). For the programmatic approach, although most studies ([Behar-Horenstein & Niu, 2011](#)) that employed this method found increases in the level of CT, the transformation occurred over a long period, during which many other factors (e.g., out-class factors) may have influenced the results. For the instructional approach, [Ennis \(1989\)](#) adopted the typology of instructional approaches, that is, critical thinking can be taught separately from subjects using (a) the general approach, be infused explicitly in instruction in existing subject matter areas using (b) the infusion approach, result from a student’s implicit immersion in the subject matter called (c) the immersion approach, or be taught as a combination of the general approach with infusion or immersion called (d) mixed approach ([Ennis, 1989](#)). However, [Abrami et al. \(2008\)](#) put forward evidence that "improvement in students’ critical thinking skills and dispositions cannot be a matter of implicit expectation" (p. 1121).

Except for the effects of cognitive, meta-cognitive, and in-class factors, others, such as demographic and out-of-class factors, can also influence CT development. Many forms of background information factors have been observed, such as gender, age, grade, academic achievement, level of education of parents, socio-economic status, residency (rural or urban areas), private or public schools, and majors. Out-of-class factors, such as maturation and significant life transitions ([Thompson & Rebesch, 1999](#)), can also influence CT. However, most studies focused on the effectiveness of teaching methods instead of these environmental factors.

1.4. PBL and its intervention on CT for higher education

As a student-centered instructional approach, PBL mainly directs students’ involvement in group study to solve ill-defined and open-ended problems using the following learning steps: analyzing problems, setting goals, collecting resources, summarizing ideas, and reflecting on problem-solving experiences ([Lin et al., 2010](#)). This process is designed to promote analytic reasoning, problem-solving, and collaborative learning, which are components of CT. Thus, theoretically, PBL is regarded as a possible practical approach for developing CT due to the overlapping contents between PBL and CT.

PBL can be categorized into two kinds based on learning places. (a) The scenario-PBL ([Lave & Wenger, 1991](#)) has a strict

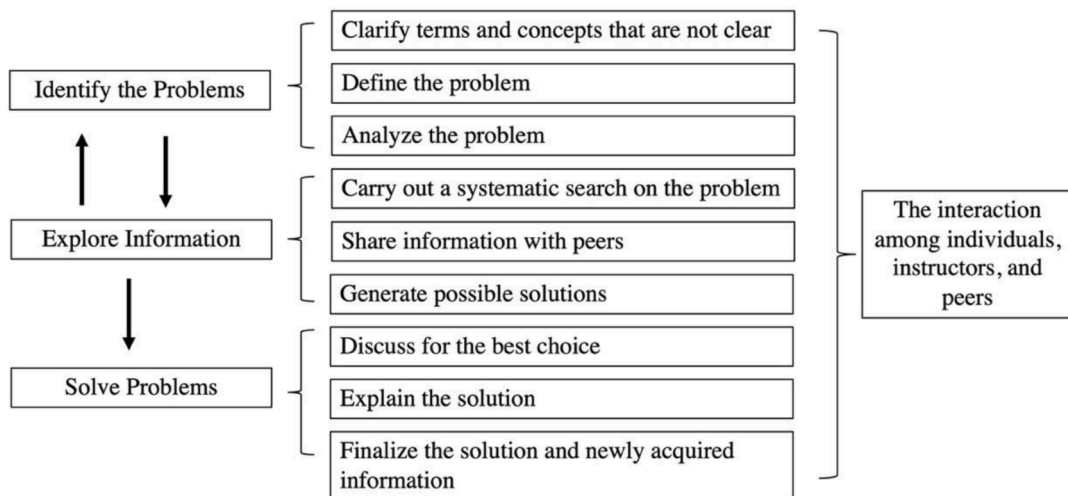


Fig. 1. The Steps of PBL.

requirement that the problem should be from real life, carried out in a real-life situation, and solved eventually. (b) The case-PBL (Bruner, 2002) is a story generated from real life to be structured into a script for students to discuss in class. In addition, based on the interaction, PBL can be categorized into collaborative PBL and individual PBL (Suebunukarn & Haddawy, 2006). However, for all studies in this meta-analysis, they adopted case-PBL and collaborative PBL since they carried out the intervention in class with students' interactions. Thus, they were named PBL generally in this study.

Besides, the characteristics of PBL in these studies are based on the summary of Savery (2015): Learning guidance mostly by students; open inquiry of authentic cases; problems related to subjects; collaboration; reflection; teachers' instruction. Moreover, the similarly fixed implementation steps of PBL in the studies in this meta-analysis can be generally listed in Fig. 1. The authentic cases were demonstrated to students, firstly, to identify the problem, and then to do an informative exploration of it. During this process, students may strengthen their knowledge of problem identification (the first step). Based on the possible solutions generated from synthesizing information, students will have a collaborative discussion to select the best answer and present it to peers and the instructor. According to the feedback, the final solution will be identified with newly acquired knowledge. Meanwhile, interaction among individuals, peers, and instructors exists in the whole procedure.

Previous studies on the effect of PBL on CT in higher education demonstrated different results. Many of these studies found a positive impact of PBL intervention on the development of CT skills and disposition (Ding, 2016; Gholami et al., 2016; Mandeville & Stoner, 2015; Martyn et al., 2014; Muehlenkamp et al., 2015; Nargundkar et al., 2014; Semerci, 2006; Sendag and Odabasi, 2009; Son, 2020; Sun et al., 2013; Tiwari et al., 2006; Ulger, 2018; Wang, 2018; Weissinger, 2003; Yu et al., 2013; Yuan et al., 2008).

However, the adverse effects are as follows. Iwaoka et al. (2010) conducted a series of interventions from 2001 to 2008 and found that two out of eight years witnessed significant gains for CT. The lack of such gains from the other years was because the instrument used cannot assess constructive tasks, such as open-ended, take-home exams, and composition of laboratory reports. Nevertheless, these tasks exerted strong positive effects on CT. One treatment exhibited no significant increase due to data attrition (Hesterberg, 2005). This result indicates that long-term treatment may lead to the withdrawal of participants and cause the consumption of students' knowledge. Pardamean (2012) conducted an intervention for dental students, which revealed no significant increase for CT due to the small sample size, selection bias, and the internal validity of the instrument. Choi et al. (2014) explained that the negative results obtained by their study might be due to the small sample size and non-randomized design. Other reasons for the weak significance are culture shock among participants, lack of time for learning, and no prior PBL experience (Sanderson, 2008).

1.5. Meta-analysis of the effect of PBL on CT in higher education

Several meta-analyses were also conducted to gain a general picture of the effect of PBL instructional intervention on CT in higher education. Kong et al. (2014) analyzed nine studies (from 1965 to 2012) on the impact of PBL on the CT of nursing students. The authors concluded that PBL could promote CT in nursing students with CCTDI obtaining a better magnitude of the effect size, whereas CCTST and W-GCTA were inconclusive. Oliveira et al. (2016) synthesized four studies whose results exhibited a slight positive effect size. In addition, Lee et al. (2016) focused on eight studies conducted from 2001 to 2014 and demonstrated that PBL was ineffective on CT. Yuan (2007) selected four studies that displayed contrasting results: effective and ineffective, whereas W-GCTA as an assessment tool was ineffective. Wang et al. (2009) used 11 articles to summarize that PBL was superior to traditional methods in the training for CT. Cheng et al. (2014) presented nine studies, which indicated that PBL was effective for CT among nursing students. Conversely, Liu et al. (2020) conducted a detailed meta-analysis by synthesizing 31 effect sizes from studies published before 2018, revealing that PBL is effective for CT skills and disposition. No difference exists for each discipline. Moreover, PBL was ineffective for nursing students but enormously effective for engineering students; the intervention of three to six months is considered a suitable duration, adoption of instructional scaffolds is suggested, and the group size is preferably four to eight.

In summary, previous studies and meta-analyses exhibited different levels of effectiveness of PBL instructional intervention on CT, which may be due to the small samples, publication bias, or sensibility. Thus, a meta-analysis that focuses on a relatively large quantity of novel studies is necessary to obtain convincing and supportive evidence for the effectiveness of PBL on CT in higher education and the identification of influential factors.

2. Purpose of this study

This study, analyzing 58 effect sizes extracted from 50 studies published from the year 2000 to 2021, aims to collect articles about PBL intervention on CT cultivation in the 21st century to explore whether publication bias, heterogeneity, between-group variance, overall effectiveness, and sub-group effectiveness exist in these articles to provide a relative supplement and elaborated picture for future relevant research. By synthesizing the effect sizes, we aim to answer the following research questions:

Research question 1: Can our meta-analysis identify publication bias?

Research question 2: Do the articles display evident heterogeneity?

Research question 3: What are the influencing factors of PBL intervention on CT (skills, disposition, and overall CT)?

Research question 4: What effect sizes will be revealed in each sub-group of the influencing factors?

3. Methods

3.1. Literature collection and screening

We searched major databases: Web of Science (WoS), Education Resources Information Center (ERIC), Google Scholar, Scopus, and ProQuest. Only English language studies were included in these databases. In addition, we extended our search to China National Knowledge Infrastructure (CNKI) as well to wider the spectrum of the available studies and to gain further insights into the field. Studies from this database were in Chinese. The screening procedure is listed in the PRISMA flow in Fig. 2. The keywords or descriptors used are critical thinking and problem-based learning in the title within the period restriction from 2000 to 2021. Since we plan to focus on developing the cultivation of CT during the 21st century, we set the period limitation of the previous 20 years.

Then, to identify studies adopting intervention with a control group and experiment group, we set the additional identification with the filter "control group" on abstracts of articles of each database except Google Scholar because of its lack of that function. After duplicating, 115 articles were identified.

We first filtered out the studies whose participants were not undergraduates during the screening phase. Then an intensive reading was conducted for excluding articles without necessary data that are the number of the experiment (EN) and control (CN) groups, their mean (EM and CM, respectively), and standard deviation (SD) for calculating effect sizes.

Finally, 50 studies fulfilled the requirements. Based on studies focusing on CT skills, CT disposition, or gender, we yielded 58 effect sizes with 5210 participants from the 50 qualified studies (see in the Appendix).

3.2. Literature coding and data input

Coding the selected articles intend to explore the influencing factors that may exert different, positive, or negative effects on PBL intervention for CT promotion. Two raters coded the 50 studies into 11 groups, as shown in Table 1.

The coding results (shown in the Appendix) were then transformed into binary data upon agreement between the two raters. The result of the kappa agreement test is 0.9664 (96.91%; Table 2), which is considered acceptable and a high-level agreement. However, the coding result should be totally in agreement. Thus, the coding results were rechecked. The finding indicated disagreements in terms of the participants' nationalities and disciplines. They are: (a) one rater regarded Turkey as an eastern country; and (b) one rater considered majors in food and sports under Arts and Sciences, respectively. After a consensus, Turkey was categorized as a western country; food and sports majors were, respectively, grouped under science and Arts since the disciplines of food and sports in the

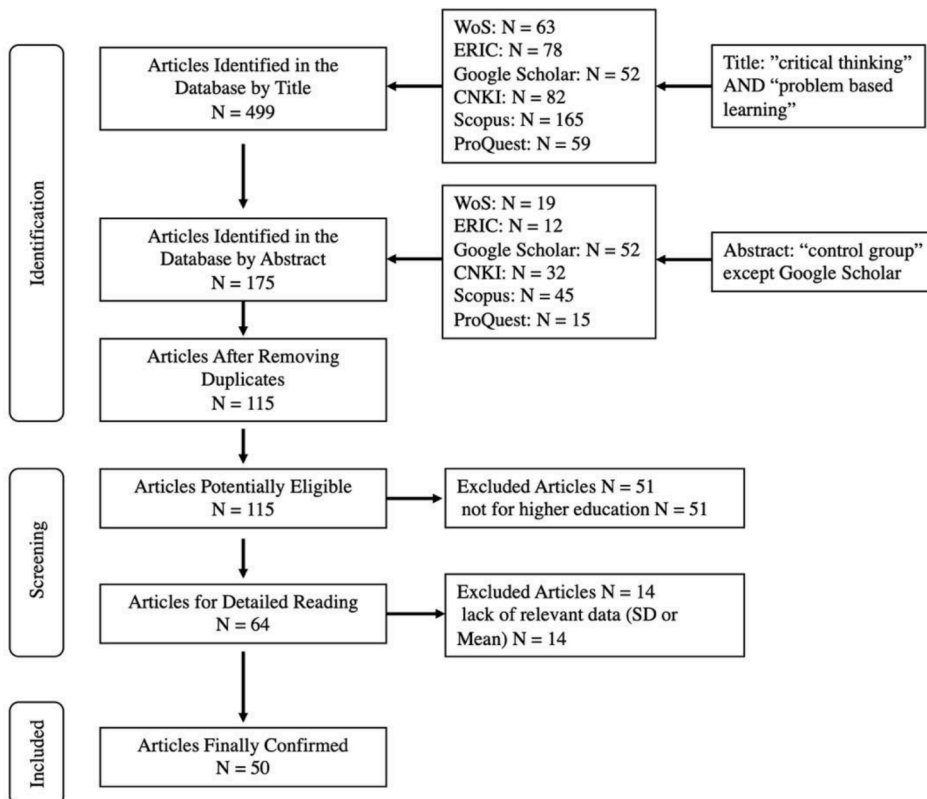


Fig. 2. PRISMA Flow of Study Analysis in Phases.

Table 1
Coding Groups of Studies.

Group	N1	N2	N3	Sub-group
Intervention Focus	27			CT skills
	31			CT disposition
Gender			1	male
			1	female
Sample Size	16	7	9	Small (< 51)
	24	13	11	Medium (= 51 - 100)
	18	7	11	Big (> 100)
Maturity	21	9	12	Juniors (first- and second-year students)
	32	15	17	Seniors (third, fourth, and higher-year students)
Instrument			28	CCTDI
		4		CCTST
		4		W-GCTA
		19	3	Others
Nationality	46	18	28	Easterners
	12	9	3	Westerners
Discipline	42	16	26	Medical Science
	10	6	4	Natural Sciences
	6	5	1	Humanity, Arts, and Social Sciences
Duration	5	1	4	Short (one to four weeks)
	12	6	6	Medium (nearly half a semester)
	41	20	21	Long (equal to or more than one semester)
Sample Type	37	15	22	Randomized
	21	12	9	Non-Randomized
Instruction Type	5	4	1	Online
	53	23	30	Offline
Group Size	10	6	4	< 6
	20	9	11	6 - 10
	5	4	1	> 10

Notes: N1, N2, and N3 represent the numbers of effect sizes in overall CT, CT skills, and CT disposition analysis, respectively. The abbreviations (Medical, Sciences, and Arts) will be used in the following texts to represent Medical Sciences, Natural Sciences, and Humanity, Arts, and Social Sciences, respectively.

Table 2
Kappa Agreement Test for Coding.

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob > Z
96.91%	8.24%	0.9664	0.0232	41.69	<0.001

studies included in this meta-analysis are more related to chemistry and social sciences, respectively. Finally, it rendered the agreement 100%. After coding, the categorized data were input into Excel for analysis.

3.3. Statistical analysis techniques

Publication bias, a measure of potential tendency in the estimated mean effect size due to publishing possibilities, aims to avoid overestimating SMD by synthesizing studies (Niu et al., 2013). Various methods can be used to verify publication bias, such as the Peters approach, Harbord approach, or Fail-Safe N (Harbord et al., 2009; Rosenthal, 1979). The variables in our study are continuous. Thus, two types of publication bias are suitable for this study, namely, (a) pictorial meta-funnel and Egger's publication bias plot, and (b) the statistical Begg's and Egger's tests.

Heterogeneity represents the consistency between all true effect sizes of the selected articles. Hedges and Olkin (2014) suggested that the primary test for effect size heterogeneity is Q statistics. However, Huedo-Medina et al. (2006) advised adopting a combinatory analysis consisting of the pictorial Galbraith plot, statistical Q, and I-square, producing more accurate results. I-square can output the proportional percentages, instead of the sampling errors, of the variability of the estimated effect sizes due to heterogeneity (Abrami et al., 2008). Moreover, additional meta-regression is a method for exploring influencing factors. Thus, this study considered the four abovementioned methods to evaluate heterogeneity.

A meta-analysis can describe the overall trend and prediction of recent, relevant empirical studies based on quantitative results. However, inconsistent values or contexts from different studies make meta-analysis cautious for researchers to conduct. The meta-analysis sensibility test intends to verify the stability of the model, the believability of the selected studies, and the analysis of the influence of the selected studies. Thus, to obtain credible research results, the sensibility of results should be tested. Without this step, the analysis of results would be considered null.

A meta-analysis is a powerful approach for summarizing and comparing results from empirical studies (Card, 2015). A meta-analysis can collect each effect size of relevant studies to compute the combined effect size (Hedges, 1981). This process is called

Standardized Mean Difference (SMD), through which the effects from each study or sub-group will be revealed. Sub-group analysis is to analyze the significance of each effect size and the between-group variance among the sub-groups coded in the literature coding procedure to identify the influential factors of the effectiveness of PBL on CT. The sample sizes of each study in this meta-analysis are big enough. Thus, this study adopted Cohen's d approach to generate SMD via STATA software to synthesize the results.

4. Results

4.1. Publication bias

Meta-funnel plot (Fig. 3) indicates that most CT skills and disposition articles are located in the upper part within the funnel and are distributed relatively symmetrically at the two sides of the middle SMD line. In addition, in Egger's publication bias plot (Fig. 4), most studies are situated on the upside of the null line (zero-line), which is relatively symmetrically distributed at the two sides of the fitting line. Moreover, we conducted Egger's test for publication bias, which automatically generates Begg's and Egger's results. As demonstrated by the results in Table 3, Begg's and Egger's overall p -values are 0.060 and 0.231, respectively, which are more than 0.05, indicating no publication bias was found. Moreover, each test's individual p -values for CT skills and disposition reached >0.05 and bias-value < 1.96 (Begg & Mazumdar, 1994). In other words, our meta-analysis contains no publication bias. Thus, the results of our study are believable, compelling, and acceptable.

4.2. Heterogeneity

Fig. 5 presents the Galbraith plot, which indicates that most of the effect sizes lie between the fitting lines (± 2 se lines), whereas others are found far from them, showing the outliers of the effect sizes where the PBL success may be over- or under-estimated for the test group. Meanwhile, the effect sizes are seemingly distributed slightly asymmetrically at the two sides of the middle linear prediction. Therefore, based on Fig. 5, the study infers that heterogeneity exists among the effect sizes, where the most influential effect sizes that may cause heterogeneity are effect sizes nos. 1, 2, 6, 11, 15, 23, 29, 30, 31, 33, 34, 38, 42, 48, 50, and 55 (see in Appendix, the second column). Moreover, for statistical evidence (Table 4), the Q statistics for CT skills and disposition are 205.00 ($I^2 = 87.3\%$) and 129.27 ($I^2 = 76.8\%$), respectively, with an overall result of 334.28 ($I^2 = 82.9\%$). Furthermore, all p -values for heterogeneity are less than 0.001. According to Higgins and Thompson (2002), I^2 values of $<30\%$ and $>70\%$ indicate slight and significant heterogeneities, respectively. These results illustrate that the selected articles display notable heterogeneity, which may be due to sub-group factors. A meta-analysis should adopt the random effect model (REM) to eliminate heterogeneity if the I^2 is $> 75\%$; otherwise, the fixed-effect model is more appropriate. Given the case of this study, REM will be employed for subsequent analysis.

To explore the potential factors that may cause heterogeneity, we conducted a meta-regression of CT skills and disposition covariates. The results for CT skills reveal that all p -values exceed 0.05. In contrast, those for CT disposition demonstrate that all p -values also exceed 0.05 except for studies that adopted medium intervention duration and online intervention (Table 5). This finding indicates that the studies on CT skills seemingly do not significantly affect the heterogeneity results. In contrast, studies that adopted medium PBL intervention duration and online intervention for CT disposition are the primary sources of heterogeneity and influenced the effect size most.

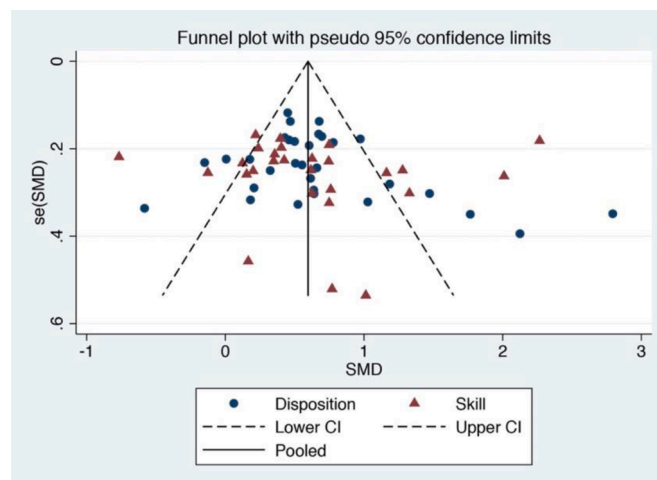


Fig. 3. Funnel Plot of the Selected Articles Varied by CT Skills and Disposition.

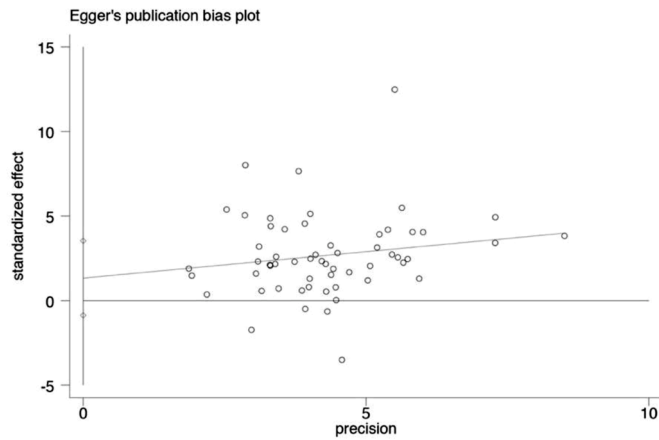


Fig. 4. Egger's Publication Bias Plot.

Table 3
Begg's and Egger's Tests for Publication Bias.

Focus	N	Begg's Score	SD	Begg's z	p	Cont. z	Corr. p	Egger's Bias	p
Skills	27	71	47.969	1.48	0.139	1.46	0.144	0.58	0.800
Disposition	31	83	58.836	1.41	0.158	1.39	0.0163	1.82	0.132
Overall	58	154	75.912	1.88	0.060	1.90	0.058	1.21	0.231

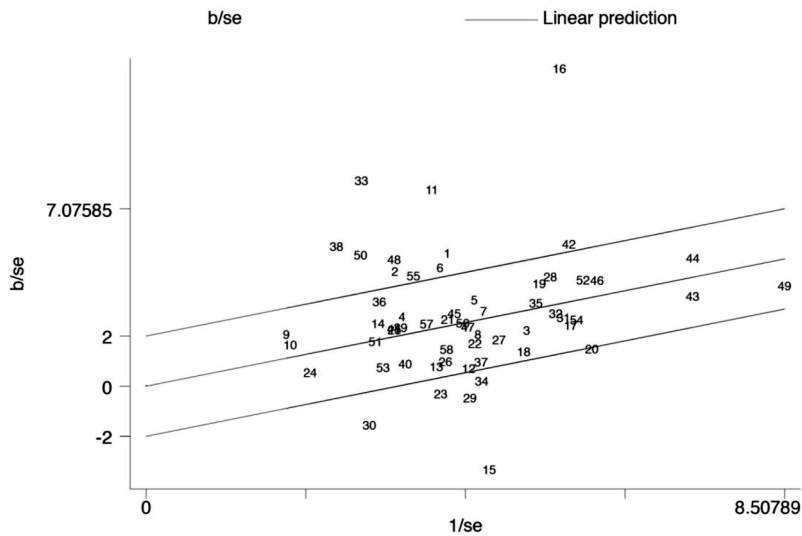


Fig. 5. Galbraith Plot for Heterogeneity.

Table 4
Statistical Overview of Overall Mean Effect Size and Test for Heterogeneity.

Model	Random	ES	N	SMD	95% CI		Heterogeneity			Sig. of SMD		
					Low	Up	Q	df	p	I ²	z	p
Skills		0.617	27	0.360	0.360	0.874	205.00	26	<0.001	87.3%	4.71	<0.001
Disposition		0.651	31	0.489	0.489	0.812	129.27	30	<0.001	76.8%	7.90	<0.001
Overall		0.640	58	0.497	0.497	0.783	334.28	57	<0.001	82.9%	8.77	<0.001

Note. ES = Effect Size; CI = Confidence Interval

Table 5
Meta-Regression of Covariates.

Group	Sub-Group	Coef.	Std. Err.	t	$p > t $
CT Disposition					
Duration	Medium	1.328	0.452	2.94	0.010
Instruction Type	Online	1.737	0.775	2.24	0.041

4.3. Sensibility

A trim-and-fill analysis was conducted to explore the sensibility of our study, in which the first result in Table 6 denotes the stability of the selected model. As the table indicates, the p -values of the fixed and random models are less than 0.001, which suggests they both work well. Moreover, the estimated SMDs are nearly the same without between-group variance and are within the valid 95% confidence interval (CI). In terms of 83.1% heterogeneity and without apparent differences between the two models, the random model is suitable for the study, which will render the results dependable.

Another output from the trim-and-fill analysis is the believability of the selected studies. Table 7 demonstrates the result of the REM with a linear trimming estimator. We find the trimming procedure with two iterations, each indicating no study requires trimming and, finally, the difference is zero. Thus, this study concluded that the selected articles are distributed relatively symmetrically and that the research results presented by these articles are believable. In addition, Figs. 6 and 7 reveal that, although several studies are situated away from the middle line, the overall figure is symmetrical. Moreover, Fig 6 presents no square sign, whereas Fig. 7 contains no filled studies. Furthermore, the effect sizes are balanced at the center point, indicating no need for articles trimmed and filled.

To explore the influence of each article on the total results, we conducted a meta-influence analysis, which aims to formulate a hypothesis by omitting only one of the effect sizes to obtain the overall SMD and corresponding CI. Fig. 8 reveals that the middle line 0.62 is the overall SMD for CT skills before omission. The left and right lines are 0.36 and 0.87, respectively, which denote the lower and upper intervals of the overall CI before deletion. Each horizontal interval indicates the CI, whereas each circle on each horizontal line represents the overall SMD after omitting the corresponding effect size along the y-axis. Fig. 9 applies the same interpretation for CT disposition. For instance, if we omit effect size no. 15, the overall SMD for CT skills will highly increase. In contrast, if effect size no. 16 is omitted, then the overall SMD for CT skills sharply decreases. Thus, we can conclude that specific effect sizes, such as nos. 11, 15, and 16 may significantly influence CT skills, whereas nos. 33, 38, 48, and 50 may largely influence the results for CT disposition. However, most SMDs tend to be stable; although they fluctuate, their values lie between the original CIs. In summary, the influence of the effect sizes from the selected articles is small, and the output results are significant and acceptable.

4.4. Results of overall and sub-group analyses

4.4.1. Overall results

The forest plot illustrates the overall distribution of effect sizes (Figure 10). The left column lists the IDs of the selected studies divided by CT skills and disposition; the zero-line is the null line; and the right column presents the SMDs and weights.

We find that the CIs of the majority of effect sizes lie at the right side of the null line, which indicates that the treatments for the experiment groups are effective for the majority of studies. In contrast, those that lie on the left side of the null line pertain to the lack of effectiveness of PBL intervention on CT. Combining the statistical results in Table 8 with those in Fig. 10, the effect sizes of CT skills and disposition are 0.617 ($p < 0.001$) and 0.651 ($p < 0.001$), respectively, with an overall effect size of 0.640 ($p < 0.001$). Furthermore, we found no between-group variance between CT skills and disposition ($p > 0.05$), demonstrating that the treatment positively affected CT skills and disposition without any difference.

The results of the sub-group comparisons on the overall CT level in Table 8 reveal the different effects of the latent influencing factors of PBL intervention on CT. For sample size, small (SMD = 0.743), big (SMD = 0.643), and medium (SMD = 0.593) sample sizes are considered effective ($p < 0.001$) with no between-group variance observed in this sub-group ($p > 0.05$). This finding indicates that sample size does not influence the degree of effectiveness.

For sample type, regardless of the assignment of the participants (random or non-random), both types exert statistical significance (p -values < 0.001); however, a between-group variance was found ($p < 0.05$). This outcome demonstrates that conducting studies on randomized samples (SMD = 0.707) is preferable to those with non-randomized samples (SMD = 0.529).

For gender, there is only one study containing gender comparison in CT disposition, and for adopted instruments, each of them focuses on either CT skills or disposition. Thus, we will show their results in the following correspondent section.

In terms of students' maturity, statistical significance ($p < 0.001$) was noted for juniors and seniors with an evident between-group

Table 6
Comparison of Results of the Two Models (Before-Trim and After-Filled).

Method	Pooled Est	95% CI		Asymptotic z	p	N	Between-Group Variance
		Lower	Upper				
Fixed	0.597	0.539	0.654	20.399	<0.001	58	0.243
Random	0.640	0.497	0.783	8.774	<0.001	58	

Table 7
Random-Effects Model with a Linear Trimming Estimator.

Iteration	Estimate	Tn	# To Trim	Diff
1	0.640	742	0	1711
2	0.640	742	0	0

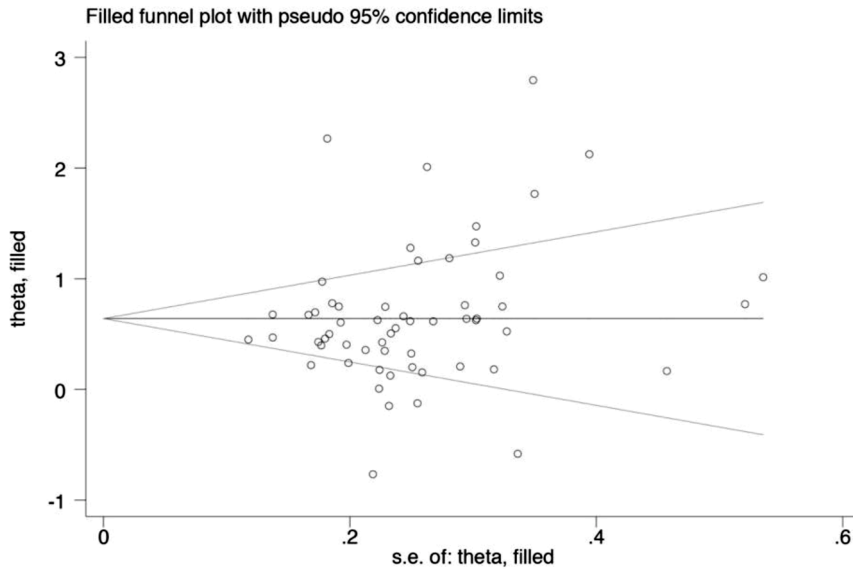


Fig. 6. Filled Funnel Plot of Trim-and-Fill Analysis.

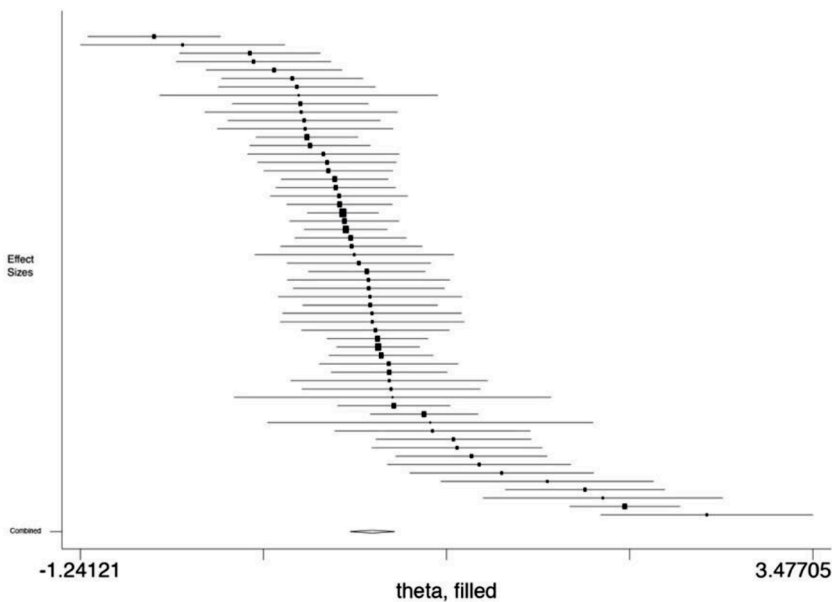


Fig. 7. Filled Forest Plot of Trim-and-Fill Analysis.

variance ($p < 0.001$), which indicates that the treatment for seniors (SMD = 0.753) is more effective than that for juniors (SMD = 0.437).

For instruction type, online and offline methods are found effective ($p < 0.001$) with between-group variance ($p < 0.05$), which indicates that online methods (SMD = 0.935) are seemingly more effective than offline traditional classroom instruction (SMD = 0.612).

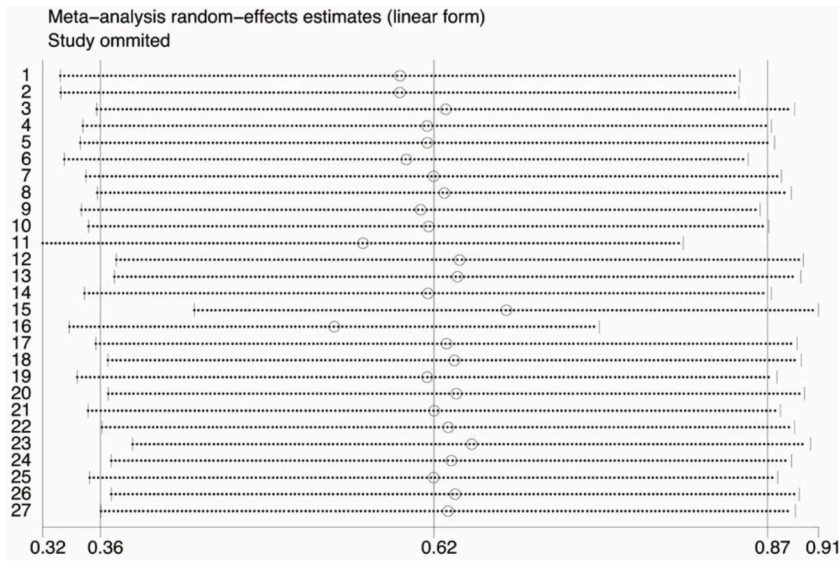


Fig. 8. Meta-Influence Analysis for CT Skills.

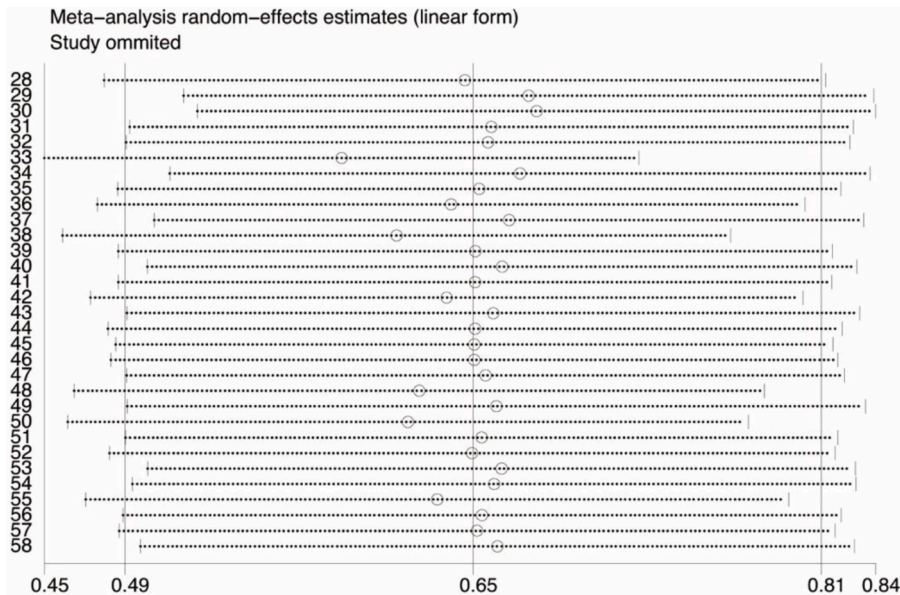


Fig. 9. Meta-Influence Analysis for CT Disposition.

According to nationality, the treatment for students from eastern or western countries (both p -values < 0.01) displayed statistical significance. In other words, the treatment is effective for eastern and western students. However, between-group variance ($p < 0.05$) was noted, which indicates that the effect size for western students ($SMD = 0.673$) is deemed to be slightly higher than that for eastern students ($SMD = 0.621$). In addition, a slight probability exists that the treatment is more effective for western students in terms of overall CT progress.

For students' disciplines, statistical significance was observed for all kinds of majors (all p -values < 0.001) without between-group variance ($p > 0.05$), which reveals that the interventions for students in all subjects are highly effective.

According to the group size, the three groups obtained significantly effectiveness ($p < 0.05$) with between-group variance ($p < 0.001$). This finding suggests that the group size of less than six ($SMD = 0.597$) is the most effective.

For treatment duration, no significance was observed for SMD for short-term interventions ($p > 0.05$), whereas medium- ($SMD = 0.786$) and long-term ($SMD = 0.619$) interventions are statistically significant (both $ps < 0.001$) without between-group variance ($p > 0.05$). These results illustrate that short-term PBL intervention exerts no effect on CT promotion. Alternatively, the effect of duration does not influence the degree of effectiveness.

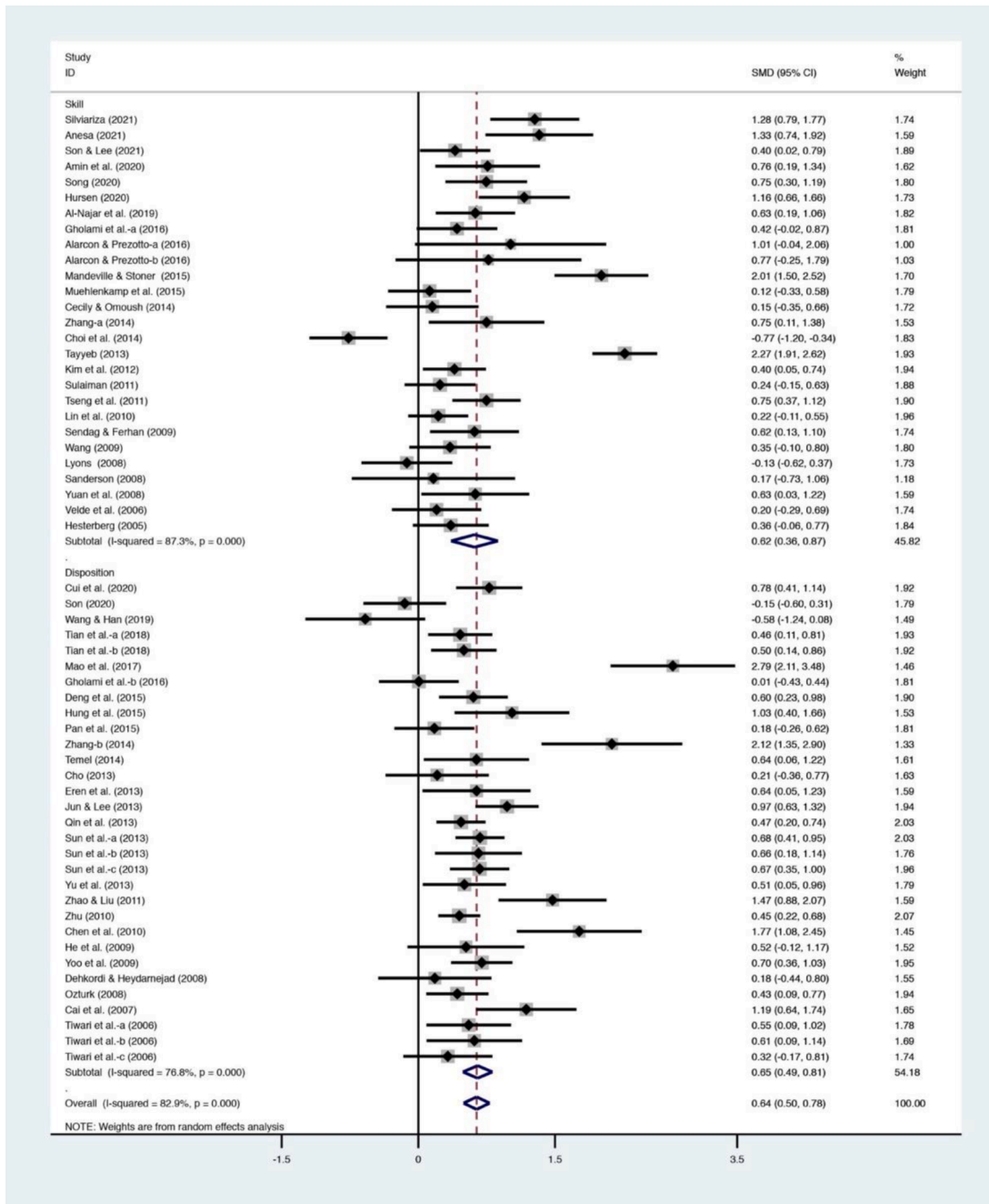


Fig. 10. Forest Plot of the Overall Mean Effect Size and Test for Heterogeneity.

4.4.2. CT skills results

Table 9 depicts the results of the sub-group comparisons of CT skills, which suggests the statistical significance (all p -values < 0.05) of the effect sizes of sample size, adopted instruments, participants' nationality, discipline, and intervention duration, all of which are with between-group variances (all p -values < 0.05).

Thus, the study infers that a big sample size ($SMD = 0.749$) is better than a small one ($SMD = 0.681$) for assessing the influence of PBL intervention on CT skills. However, a medium sample size ($SMD = 0.498$) lies behind both.

For the adopted instruments, W-GCTA ($SMD = 0.369$) and CCTST ($SMD = 0.374$) displayed weak effects in assessing progress in CT, whereas other less popular or developed instruments ($SMD = 0.721$) exerted significant effects.

For nationalities, the overall result indicates a slight probability that the treatment is more effective for western students ($SMD = 0.706$) in terms of CT skills than eastern students ($SMD = 0.552$).

Table 8
Mean Effect Sizes of Overall CT for Sub-groups.

	Effect N	SMD	95% CI Low	Up	Sig. of SMD z	p	Between-Group Variance p
Focus							
Skills	27	0.617	0.360	0.874	4.71	<0.001	0.920
Disposition	31	0.651	0.489	0.812	7.90	<0.001	
Sample Size							
Big	18	0.643	0.445	0.841	6.37	<0.001	0.061
Medium	24	0.593	0.325	0.861	4.34	<0.001	
Small	16	0.743	0.429	1.040	4.71	<0.001	
Sample Type							
Randomized	37	0.707	0.515	0.898	7.23	<0.001	.010
Non-Randomized	21	0.529	0.320	0.739	4.95	<0.001	
Maturity							
Junior	21	0.437	0.258	0.615	4.80	<0.001	<0.001
Senior	32	0.753	0.534	0.973	6.73	<0.001	
Instruction Type							
Offline	53	0.612	0.462	0.762	8.01	<0.001	0.040
Online	5	0.935	0.451	1.428	3.78	<0.001	
Nationality							
Eastern	46	0.621	0.481	0.760	8.73	<0.001	0.010
Western	12	0.673	0.211	1.135	2.86	.004	
Discipline							
Medical	42	0.580	0.412	0.749	6.76	<0.001	0.062
Sciences	10	0.740	0.376	1.103	3.99	<0.001	
Arts	6	0.907	0.481	1.334	4.17	<0.001	
Group Size							
<6	10	0.597	0.071	1.122	2.23	.026	<0.001
6–10	20	0.501	0.351	0.651	6.55	<0.001	
>10	5	0.565	0.362	0.769	5.45	<0.001	
Duration							
Long	41	0.619	0.456	0.782	7.45	<0.001	0.326
Medium	12	0.786	0.455	1.117	4.66	<0.001	
Short	5	0.451	-0.311	1.213	1.16	0.246	

For discipline, the PBL treatment on CT skills for students in Sciences (SMD = 0.869) is the most effective, whereas those for arts and medical students are lagging. Intervention for students in the Arts (SMD = 0.707) is better than those in Medicine (SMD = 0.492).

Only one study conducted a short-term intervention; we cannot discuss it. Medium-term interventions (SMD = 0.561) reveal less significant effectiveness than long-term ones (SMD = 0.586).

Although we noted no significance ($p > 0.05$) between the treatments for juniors and groups with less than six members. A strong effect was observed for seniors (SMD = 0.769, SMD p -value < 0.001), and a group with more than 10 members (SMD = 0.578, $p < 0.01$) that is slightly more effective than groups with 6 to 10 members (SMD = 0.480, $p < 0.001$) with significant between-group variance ($p < 0.001$).

Furthermore, the effect sizes were non-significant for the selected studies with randomized and non-randomized samples and online and offline instruction with no significant between-group variance ($p > 0.05$). Nevertheless, the treatments remain effective ($p < 0.01$) with SMDs of 0.647 and 0.594 for randomized and non-randomized samples, respectively. Lastly, the SMDs for online and offline instruction are 0.809 and 0.581, respectively.

4.4.3. CT disposition results

Table 10 depicts the results of the sub-group comparisons for CT disposition. Only one study focused on gender differences; thus, we cannot obtain accurate results. However, we opted to calculate the effect size of the said study to provide a general indication that the effects of PBL intervention for CT among female (ES = 0.673, $p < 0.001$) and male (ES = 0.660, $p < 0.01$) are positive without between-group variance (p -value > 0.05).

Small (SMD = 0.708), medium (SMD = 0.709), and big (SMD = 0.593) sample sizes are deemed effective ($p < 0.01$) without between-group variance ($p > 0.05$).

Studies using randomized samples (SMD = 0.750, $p < 0.001$) are more effective than those adopting non-randomized samples (SMD = 0.481, $p < 0.001$) with between-group variance ($p < 0.05$).

For maturity, the intervention on CT disposition is effective ($p < 0.001$) for seniors and juniors (SMDs: 0.733 and 0.540, respectively) without between-group variance ($p > 0.05$).

Except for CCTDI (SMD = 0.676, $p < 0.001$), other types of instruments ($p > .05$) are deemed ineffective for PBL intervention on CT disposition.

In contrast to CT skills, the effect on eastern students (SMD = 0.663, $p < 0.001$) displayed no between-group variance ($p > 0.05$) from western students (SMD = 0.514, $p < .001$).

For instruction type, studies that used online media (SMD = 1.473, $p < 0.001$) are more effective than those that adopted an offline

Table 9
Mean Effect Sizes of CT Skills for Sub-groups.

	Effect N	SMD	95% CI		Sig. of SMD		Between-group Variance <i>p</i>
			Low	Up	<i>z</i>	<i>p</i>	
Sample Size							
Big	7	0.749	0.098	1.399	2.45	0.014	0.011
Medium	13	0.498	0.135	0.860	2.66	0.008	
Small	7	0.681	0.385	0.976	6.01	<0.001	
Sample Type							
Randomized	15	0.637	0.287	0.986	3.57	<0.001	0.146
Non-Randomized	12	0.594	.200	0.988	2.95	0.003	
Maturity							
Junior	9	0.270	-0.101	0.640	1.43	0.153	<0.001
Senior	15	0.769	0.405	1.133	4.14	<0.001	
Instruction Type							
Offline	23	0.581	0.290	0.873	3.91	<0.001	0.315
Online	4	0.809	0.302	1.317	3.13	0.002	
Instrument							
CCTST	4	0.374	0.100	0.648	2.67	0.008	0.007
W-GCTA	4	0.369	0.154	0.585	3.36	0.001	
Others	19	0.721	0.369	1.074	4.01	<0.001	
Nationality							
Eastern	18	0.552	0.322	0.783	4.69	<0.001	<0.001
Western	9	0.706	0.090	1.321	2.25	0.025	
Discipline							
Medical	16	0.492	0.138	0.845	2.73	0.006	0.032
Sciences	6	0.869	0.222	1.515	2.63	0.008	
Arts	5	0.707	0.424	0.990	4.90	<0.001	
Group Size							
<6	6	0.486	-0.238	1.209	1.32	0.188	<0.001
6-10	9	0.480	0.205	0.756	3.42	0.001	
>10	4	0.578	0.345	0.811	4.87	<0.001	
Duration							
Long	20	0.586	0.251	0.920	3.43	0.001	0.014
Medium	6	0.561	0.275	0.847	3.85	<0.001	
Short	1	1.278	0.790	1.767	5.13	<0.001	

approach (SMD = 0.626, $p < 0.001$) with between-group variance ($p < 0.01$). Notably, however, only one study used the online approach for disposition treatment.

The treatment for students in Arts (SMD = 2.125) exhibited stronger effect than that for students in Medicine (SMD = 0.628) and Sciences (SMD = 0.552) with between-group variance (p -value < 0.01). However, only one study assessed the effectiveness of PBL intervention on CT disposition among students in Arts.

For the group size, only one study focused on groups with more than ten members (SMD = 0.506, $p < 0.05$). Conversely, groups with less than six members displayed no significance in terms of intervention ($p > 0.05$). However, treatments for groups with 6 to 10 are effective (SMD = 0.517, $p < 0.001$).

A short-term intervention remains non-significant ($p > 0.05$) but with between-group variance ($p < 0.001$), which indicates that medium-term intervention (SMD = 0.988, $p < 0.01$) is better than long-term intervention (SMD = 0.613, $p < 0.001$).

5. Discussions, limitations, and future directions

5.1. Discussions

5.1.1. Technical results

The study aims to explicate the sources of heterogeneity in the meta-analysis. Combining the results of the Galbraith plot, meta-regression, meta-influence, and forest plot, we find that the specific studies, effect size, or sub-group factors, such as effect size nos. 11, 15, 16, 30, 33, 38, 48, and 50, and studies that adopted online instructional methods and medium-term intervention for CT disposition, may cause heterogeneity, which illustrates these two can enhance the effectiveness of PBL on CT. Based on these studies, we identify three commonalities as follows: (a) adoption of instruments lacking validity from large-scale testing (Cai et al., 2007; Choi et al., 2014; Jun & Lee, 2013; Lyons, 2008; Mandeville & Stoner, 2015; Tayyeb, 2013; Zhang, 2014; Zhao & Liu, 2011), which may lead to extreme positive or negative results; (b) outcomes including other focus: the targeting construct is vital for result validity. Many studies (Choi et al., 2014; Lyons, 2008; Mandeville & Stoner, 2015; Tayyeb, 2013), including those on CT, problem-solving, evidence-based practice, effective communication, self-directed learning, content knowledge, and academic achievements, may lead to the difference; and (c) mixed input methods: some studies (Chen et al., 2010; Jun & Lee, 2013; Mao et al., 2017) combine other teaching methods, such as case-based study, and simulation as PBL. Although these methods can also be regarded as one aspect of PBL, they enlarge or decrease the effect of pure PBL. Since the studies in our meta-analysis demonstrate no publication bias, acceptable

Table 10
Mean Effect Sizes of CT Disposition for Sub-groups.

	Effect N	SMD/ES	95% CI		Sig. of SMD		Between-Group Variance
			Low	Up	z	p	p
Sample Size							
Big	11	0.593	0.495	0.691	11.91	<0.001	0.835
Medium	11	0.709	0.300	1.118	3.40	0.001	
Small	9	0.708	0.214	1.201	2.81	0.005	
Sample Type							
Randomized	22	0.750	0.521	0.978	6.42	<0.001	0.031
Non-Randomized	9	0.481	0.343	0.620	6.82	<0.001	
Gender							
Male	1	0.660	0.347	0.999	2.71	0.007	0.848
Female	1	0.673	0.487	0.780	4.04	<0.001	
Maturity							
Junior	12	0.540	0.407	0.673	7.95	<0.001	0.277
Senior	17	0.733	0.459	1.008	5.624	<0.001	
Instrument							
CCTDI	28	0.676	0.506	0.846	7.80	<0.001	0.572
Others	3	0.413	-0.239	1.066	1.24	0.215	
Nationality							
Eastern	28	0.663	0.487	0.840	7.35	<0.001	0.532
Western	3	0.514	0.250	0.777	3.82	<0.001	
Instruction Type							
Offline	30	0.626	0.466	0.786	7.66	<0.001	0.003
Online	1	1.473	0.880	2.066	4.87	<0.001	
Discipline							
Medical	26	0.628	0.451	0.806	6.94	<0.001	<0.001
Sciences	4	0.552	0.307	0.737	4.76	<0.001	
Arts	1	2.125	1.352	2.897	5.39	<0.001	
Group Size							
<6	4	0.774	-0.074	1.622	1.79	0.074	0.100
6-10	11	0.517	0.333	0.701	5.51	<0.001	
>10	1	0.506	0.048	0.963	2.17	0.030	
Duration							
Long	21	0.613	0.482	0.743	9.20	<0.001	0.001
Medium	6	0.988	0.384	1.592	3.21	0.001	
Short	4	0.238	-0.579	1.055	.570	0.568	

heterogeneity, and sensibility with a stable model, revealing the general trend of all objectively existing studies that should be analyzed, we cannot omit them. Despite the omission, heterogeneity I^2 remains at 42.3%.

5.1.2. Research results

5.1.2.1. Results on overall CT. The overall effect size of this meta-analysis, similar to the meta-analysis of Liu et al. (2020), which includes 31 studies, indicates that PBL interventions on CT (skills and disposition) are effective. However, the selected studies did not discuss between-group variance. They concluded that the effectiveness of PBL on CT disposition is more significant than that on CT skills, which is not dependable. Thus, we conducted a variance comparison and determined that no between-group variance exists, indicating that PBL intervention for CT is effective, and the degree of influence on skills and disposition is nearly the same.

Besides, we concluded that sample size, discipline, and treatment duration do not affect overall CT. However, previous studies obtained inconsistent results, such as the empirical review of Masek and Yamin (2011). Choi et al. (2014) summarized that a small sample size might influence the results, whereas Kong et al. (2014) concluded the opposite for nursing students. The results of Liu et al. (2020) indicated ineffectiveness, which is in line with the current study in terms of duration. This notion suggests that, although there is no between-group difference, the medium is suitable, whereas Masek and Yamin (2011) argued that long-term intervention is better.

In this study, the potential factors are participants' maturity, nationality, instruction type, group size, and types of sample selection. Although we could not find studies that focused on maturity, the cultivation of CT among senior students may be more effective than that for junior students. The plausible explanation for the difference is that with life experience increasing, seniors may acquire more practice in thinking and problem-solving knowledge. Besides, seniors may face the choice of life directions identifying a job or higher degree, which forces them to think critically.

Regarding nationality and discipline, we obtained results that differed from the meta-analysis of Leng and Lu (2020), which explored the overall teaching effectiveness, not specific instructional methods on CT. However, our study only focused on PBL. Thus, the difference is considered acceptable. Besides, CT courses or assessments in eastern countries are not regarded as a standard method embedded in their curricula (Leng & Lu, 2020; Li & Pan, 2019). Moreover, most instruments used for CT assessment are textual, as CT is related to reading skills (Csapó & Csépe, 2012), which may underlie the better performance of students in arts than that of students in other disciplines. However, this aspect requires further exploration.

Furthermore, randomized samples are regarded as more objective than non-randomized ones. The literature does not provide this

conclusion, whereas a few non-randomized studies (e.g., Çeliker, 2021) listed it as a limitation. Nevertheless, the online instructional approach, as a new and attractive method, may arouse more interest in students than the traditional classroom methods that students may be bored with. Group size determines the interaction among peers. All group sizes are effective, but those less than 6 and more than 10 witnessed a little higher effect size than the medium one, which may be caused by the degree of interaction. For small group sizes, students can have more time for them to think independently to enhance CT individually. For big group size, students can have more collaboration with peers, by which, they can obtain more training on thinking and evaluating. While for medium group size, they have neither such amount of time for individual thinking nor more interactions.

5.1.2.2. Results on CT skills. For CT skills, no difference was observed for the sample type, and providing intervention for students online or offline is not an influencing factor in CT skills cultivation. Besides, studies on this aspect are lacking. Nevertheless, other factors are apparent.

For maturity and nationality, several studies conducted on the effectiveness of teaching CT (Leng & Lu, 2020) obtained results similar to those of the current study, and the plausible explanations for the larger effect size for seniors and westerners have been shown in the discussion of the results of overall CT.

For instruments, widely-used two are deemed effective but lower than others. Niu et al. (2013) also presented this consequence, which may be the reason that widely-used ones are standardized and require strict conditions to use.

For discipline, Liu et al. (2020) included ten studies and obtained no difference among each subject. However, the current study consists of 30 articles. In summary, we argue that the treatment for students in Sciences is slightly better than those for students in Arts and Medicine on CT skills. Students in Sciences have more training on logical thinking, which can help them acquire CT skills better, while those in Arts are good at open-thinking, with the relatively lower start point of CT skills, the good effect size can be understood. However, for medical students, their prior focus tends to be operational skills instead of thinking skills, which may be the reason for the lower effect size.

The result on the group size in this study differs from that of Liu et al. (2020). However, the previous research lacked relevant data on this issue due to insufficient studies. Thus, our results will be preliminary to demonstrate that a group with more than ten members is better for CT skills cultivation. The possible explanation is that skills need more practice for enhancement. The more students in a group, the more ideas and discussions will pump out, which can also be an illustration of the results of the sample size.

Only one study analyzed short-term intervention for CT skills. Long-term intervention is better than medium-term intervention for CT skills promotion. However, both are nearly the same, which agrees with the result of Masek and Yamin (2011) but not Liu et al. (2020). This also proves that CT skills need more chance for practicing.

5.1.2.3. Results on CT disposition. For CT disposition, sample type, instruction type, discipline, and treatment duration are the factors that influence the degree of effectiveness. The sample and instruction types results are similar to the overall CT, indicating that randomized samples and online instruction might display better results. However, there is still no previous study on them.

In terms of discipline, only one study was conducted on Arts. Thus, we did not provide a discussion on this topic. However, the results for CT disposition are in contrast with those for CT skills; that is, treatment for students in Medicine is better than those in the Sciences. This finding is inconsistent with previous studies, such as Liu et al. (2020), who proposed that no difference exists among disciplines. The plausible explanation is that CT disposition is a psychometric facet of CT, and medical students have more interaction with real lives than those in Sciences, which leads to the better acquirement of psychological disposition.

For the duration, medium-duration intervention exerted a much more significant effect than the long one, and short-term ones are considered ineffective. This result is consistent with that of Liu et al. (2020). However, the statistical results with between-group variance differed from the previous one, including 21 studies (Liu et al., 2020). Possible explanations are that compared with the results of CT skills regarding duration that long-term intervention seems more effective, as a psychological disposition, longer treatment may consume this motivation and the inclination of CT cannot form during such a short intervention.

Moreover, only one study was conducted on gender differences. Thus, we were unable to use combined effect sizes to evaluate it. Nevertheless, we calculate the effect size of gender, which indicated that no gender difference exists in terms of PBL intervention for CT skills promotion, which is consistent with previous studies (Sun et al., 2013; Weissinger, 2003).

For factors that do not exert any influence, we could not find studies on the effect of sample size and the maturity of students on CT disposition. However, no difference among sub-groups was noted. For the instruments, CCTDI is effective, whereas other instruments are ineffective despite the lack of between-group variance. This result indicates that mature tools are better than those that did not undergo large-scale testing and validation. We obtained the same results as Liu et al. (2020), who proposed that no differences existed among the three types of groups. However, groups with approximately six members are better for cultivating CT disposition.

5.2. Limitations

Although evidence to prove its dependability, the study has its limitations. The first is the categories of the sub-groups. For example, as the joint between Asia and Europe, Turkey is not consistent in belonging to western or eastern countries. Therefore, other researchers may have different sub-dimensions of these categories. Since another widely-used instrument, CCTT, is used in these studies only once, we grouped it under "others", which may underestimate its effectiveness. Moreover, we cannot include unpublished papers, and articles from all the databases in all languages, which may lead to bias theoretically. In the identification of influential factors, this study adopted the factor-by-factor approach assuming the factors' effects would not be modulated by any other. However, in reality,

they will affect each other and jointly contribute to the changes detected.

5.3. Future directions

5.3.1. Research directions

For future studies, if the scope of one meta-analysis is technically narrow, then the meta-analysis should be conducted with caution as it may lead to publication bias or instability of the research. In this case, enlarging the research scale or waiting for sufficient articles is reasonable. Meanwhile, caution should also be practiced regarding the sub-group categories. In general, PBL instructional intervention for CT is considered effective. However, future studies should conduct an in-depth investigation and may need to explore the prior CT level, out-of-class experience, various types of PBL activities (i.e., scenario-PBL), and the group study scaffolding of students and instructors. Besides, to extend this topic, future studies may focus on PBL and the connection between other or various instructional approaches and CT progress. In addition, further studies on gender may be conducted due to the lack of relevant studies currently. Lastly, the study recommends using widely-used instruments, such as CCTST and CCTDI, and avoiding tools that lack validity or reliability testing. Moreover, although instruments for assessing CT, such as CLA+, HEIghten, and HCTA, lack application for evaluating the effectiveness of PBL on CT, as they have undergone large-scale testing. Thus, they are suitable for use in future studies. Finally, to avoid the bias of factor-by-factor analysis, the future study can adopt an advanced method for the identification of influencing factors. In summary, the selected articles mainly focused on the overall and sub-dimension of CT. However, research should be conducted on each sub-dimension below CT skills and disposition, providing future CT assessment and intervention directions.

5.3.2. Instructional directions

If the aim of future instructions tends to strengthen overall CT, the intervention is better to be carried out in the class of small samples (better less than 51) with groups of less than six students. Besides, if possible, online instruction or computer-assisted instruction can be considered effective methods. Moreover, to cultivate CT, it is better to carry out this instructional activity for senior students.

If instructions aim to enhance CT skills especially, a big class (can be > 100) and big group size (> 10) can be suggested. Besides, the instructional duration can be a little longer (one semester or more). Still, starting the CT skills treatment from the third year will be better. If the intervention especially focuses on CT disposition improvement, a medium duration (half a semester) is suggested. In addition, online instruction can also be considered if it is technically allowed.

6. Conclusion

This meta-analysis, without publication bias and with acceptable stability of results, reveals that (a) PBL instructional intervention on CT (skills and disposition) is powerfully effective; (b) the sources of heterogeneity are studies with other instruments and mixed input methods and outcomes; and (c) the influencing factors that may influence the effectiveness of PBL on overall CT are students' maturity, nationality, instruction type, sample type, and group size. In terms of CT skills, these factors are sample size, maturity, instruments, nationality, discipline, group size, and treatment duration. In contrast, those for CT disposition are sample type, instruction type, discipline, and treatment duration. Moreover, despite several positive results revealed by recent studies, future in-depth studies should be conducted to investigate other aspects of PBL intervention on CT and its various dimensions by considering all kinds of research-related issues. Besides, future CT instructional interventions should consider the influential factors in to structure the class better for more effective cultivation.

Studies included in the meta-analysis

(Al-Najar et al., 2019; Alarcon & Prezotto, 2016; Amin et al., 2020; Anesa & Ahda, 2021; Cai et al., 2007; Cecily & Omoush, 2014; Chen et al., 2010; Cho, 2013; Choi et al., 2014; Cui et al., 2020; Dehkordi & Heydarnejad, 2008; Deng et al., 2015; Eren & Akinoglu, 2013; Gholami et al., 2016; He et al., 2009; Hesterberg, 2005; Hung et al., 2015; Hursen, 2020; Jun & Lee, 2013; Kim et al., 2012; Lee & Son, 2021; Lin et al., 2010; Lyons, 2008; Mandeville & Stoner, 2015; Mao et al., 2017; Muehlenkamp et al., 2015; Ozturk et al., 2008; Pan et al., 2015; Qin et al., 2013; Sanderson, 2008; Sendag & Odabasi, 2009; Silviariza et al., 2021; Silviariza et al., 2020; Son, 2020; Song, 2020; Sulaiman, 2011; Sun et al., 2013; Tayyeb, 2013; Temel, 2014; Tian et al., 2018; Tiwari et al., 2006; Tseng et al., 2011; Velde et al., 2006; Wang, 2009; Wang & Han, 2019; Yoo et al., 2009; Yu et al., 2013; Yuan et al., 2008; Zhang, 2014; Zhao & Liu, 2011; Zhu, 2010).

CRedit authorship contribution statement

Yong Liu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft. **Attila Pásztor:** Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

We have no conflicts of interest to disclose.

Acknowledgement

This is to acknowledge that this work is supported by China Scholarship Council (Grant Number: 202008500108), Tempus Public Foundation of Hungary government (Grant Number: SHE-16610-004/2020), the Research Program for Public Education Development of the Hungarian Academy of Sciences (Project Number: KOZOKT2021-16), and University of Szeged (Open Access Grant Number: 5548).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.tsc.2022.101069](https://doi.org/10.1016/j.tsc.2022.101069).

References

- 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. <https://doi.org/10.3102/0034654308326084>
- Adler, M. J. (2000). *How to think about the great ideas: From the great books of western civilization* (M. Weismann, Ed.). Open court.
- Alarcon, M. F. S., & Prezotto, K. H. (2016). Evaluation of educational strategy, grounded on problem-based learning on nursing undergraduate. *Rev Rene*, 17(2), 242–249. <https://doi.org/10.15253/2175-6783.2016000200012>
- Al-Najar, H., Khalil, A., Awa, S., Bakar, A., Siah, N., & Aziz, A. (2019). Problem-based learning (PBL) versus lecture based learning (LBL): Effect on the development of critical thinking, problem solving and self directive learning skills in nursing students. *Journal of Nursing Care*, 8(3), 489. <https://www.hilarispublisher.com/open-access/problem-based-learning-pbl-versus-lecture-based-learning-lbl-effect-on-the-development-of-critical-thinking-problem-solvi.pdf>
- Amin, S., Utaya, S., Bachri, S., & Sumarmi, S. (2020). Effect of problem based learning on critical thinking skill and environmental attitude. *Journal for the Education of Gifted Young Scientists*, 743–755. <https://doi.org/10.17478/jegys.650344>
- Anesa, D. R., & Ahda, Y. (2021). The effectiveness of problem based learning (PBL) based e-module on the classic genetic materials to improve the student's critical thinking skills. *International Journal of Social Science and Human Research*, 4(7), 1785–1789. <https://doi.org/10.47191/ijsshr/v4-i7-30>
- Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching and Learning*, 8, 25–41. <https://doi.org/10.19030/tlc.v8i2.3554>
- Bensley, D. A., & Murtagh, M. P. (2012). Guidelines for a scientific approach to critical thinking assessment. *Teaching of Psychology*, 39(1), 5–16. <https://doi.org/10.1177/0098628311430642>
- Bernard, R., Zhang, D., Abrami, P., Sicolu, F., Borokhovski, E., & Surkes, M. (2008). Exploring the structure of the Watson–Glaser critical thinking appraisal: One scale or many subscales? *Thinking Skills and Creativity*, 3, 15–22. <https://doi.org/10.1016/j.tsc.2007.11.001>
- Bruner, R. F. (2002). Socrates' Muse: Reflections on effective case discussion leadership. *Social Sciences Education eJournal*. 10.2139/ssrn.320808.
- Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., & Almeida, L. S. (2012). The Halpern critical thinking assessment and real-world outcomes: Cross-national applications. *Thinking Skills and Creativity*, 7(2), 112–121. <https://doi.org/10.1016/j.tsc.2012.04.001>
- Cai, J., Nie, H., Yu, X., & Sun, B. (2007). Yaolixue PBL jiaoxue zhong xuesheng pipanxingsiweinengli de peiyang [Evaluation of critical thinking skills in pharmacological PBL teaching]. *China Higher Medical Education*, (2), 4–5. <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=ZOGU200702002&DbName=CJFQ2007>
- Card, N. A. (2015). *Applied meta-analysis for social science research*. Guilford Publications.
- Cecily, H. S. J., & Omoush, A. (2014). Efficacy of problem Based Learning (PBL) over Lecture method in enhancing the critical thinking skills and problem solving ability among nursing students in KSA. *International Journal of Educational Science and Research (IJESR)*, 4, 1–8. <https://www.researchgate.net/publication/273772071>
- Çeliker, H. D. (2021). Problem-based scenario method with experiments: Determining the prospective science teachers' biology self-efficacy and critical thinking tendency. *Science Education International*, 32(1), 23–33. <https://doi.org/10.33828/sei.v32.i1.3>
- Chen, X., Ruan, L., & Gao, G. (2010). Jiyu CBS+PBL de huli chafang zai peiyang xuesheng pipanxingsiweinengli zhong de yingyong [Application of nursing ward rounds based on CBS and PBL in cultivating critical thinking abilities of nursing students]. *Chinese Journal of Nursing Education*, 7(9), 405–407. <https://doi.org/10.3761/j.issn.1672-9234.2010.09.008>
- Cheng, L., Cui, W., Zhao, N., Cui, X., & Chen, J. (2014). A Meta analysis of effect of problem-based learning on development of nursing students' critical thinking. *Chinese Journal of Practical Nursing*, 30(28), 67–71. <https://doi.org/10.3760/cma.j.issn.1672-7088.2014.28.020>
- Cho, H.-Y. (2013). The effect of SIM-PBL education on critical thinking disposition and self-efficacy in paramedic students. *The Korean Journal of Emergency Medical Services*, 17(3), 61–71. <https://doi.org/10.14408/KJEMS.2013.17.3.061>
- Choi, E., Lindquist, R., & Song, Y. (2014). Effects of problem-based learning vs. traditional lecture on Korean nursing students' critical thinking, problem-solving, and self-directed learning. *Nurse Education Today*, 34(1), 52–56. <https://doi.org/10.1016/j.nedt.2013.02.012>
- Cooper, J. L. (1995). Cooperative learning and critical thinking [Editorial Material]. *Teaching of Psychology*, 22(1), 7–9. https://doi.org/10.1207/s15328023top2201_2
- Csapó, B., & Csépe, V. (2012). *Framework for diagnostic assessment of reading*. Nemzeti Tankönyvkiadó. http://pedagogus.edia.hu/sites/default/files/reading_framework_english.pdf
- Cui, D., Wang, S., Zhuang, L., & Wu, X. (2020). PBL jiaoxuefa zai huli zonghe jineng xunlianke shang tigao husheng de pipanxingsiweinengli [Using PBL Teaching Method to Improve Critical Thinking Ability of College Nursing Students in the Nursing Skills Training]. *Journal of Kunming Medical University*, 41(4), 167–171. <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=KMYX202004032&DbName=CJFQ2020>
- Dehkordi, A. H., & Heydarnejad, M. S. (2008). The effects of problem-based learning and lecturing on the development of iranian nursing students' critical thinking. *Pakistan Journal of Medical Sciences*, 24(5), 740–743. <http://eprints.skums.ac.ir/6778/1/1.pdf>
- Dell'Olio, J. M., & Donk, T. (2007). *Models of teaching: Connecting students with learning standards*. Sage Publications. <https://doi.org/10.4135/9781452232324>
- Deng, W., Ren, B., Hu, C., Zhao, J., Qi, X., & Peng, F. (2015). PBL jiaoxuefa zai huli benkesheng pipanxingsiweinengli peiyang zhong de yingyong xiaoguo pingjia [Evaluation on application effect of PBL teaching method in cultivation of critical thinking ability of nursing undergraduates]. *Chinese General Practice Nursing*, 13(36), 3637–3640. <https://doi.org/10.3969/j.issn.1674-4748.2015.36.001>
- Ding, X. (2016). The effect of WeChat-assisted problem-based learning on the critical thinking disposition of EFL learners. *International Journal of Emerging Technologies in Learning (IJET)*, 11(12). <https://doi.org/10.3991/ijet.v11i12.5927>

- Dwyer, C., Hogan, M. J., & Stewart, I. (2011). The promotion of critical thinking skills through argument mapping. In C. P. Horvath, & J. M. Forte (Eds.), *Critical Thinking* (1 ed.). Nova Science Publishers <https://www.researchgate.net/publication/259286333>.
- Elder, L., & Paul, R. (2020). *Critical thinking: Tools for taking charge of your learning and your life* (third ed.). Rowman & Littlefield <https://books.google.cg/books?id=yBzZDwAAQBAJ&lpg=PP1&hl=fr&pg=PR5#v=onepage&q&f=false>.
- Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4–10. <https://doi.org/10.3102/0013189x018003004>
- Ennis, R. H. (2011). Critical thinking: Reflection and perspective. In *Inquiry: Critical Thinking Across the Disciplines*, 26, 4–18. <https://doi.org/10.5840/inquiryctnews20112613>. Prentice-Hall.
- Ennis, R.H., Gardiner, W.L., Morrow, R., Paulus, D., & Ringel, L. (1964). The cornell class-reasoning test, Form X. *Champaign: Illinois critical thinking project, department of educational policy studies, University of Illinois at Urbana-Champaign*. <http://criticalthinking.net/wp-content/uploads/2018/01/CornellConditionalReas.pdf>.
- Eren, C. D., & Akinoglu, O. (2013). Effect of problem-based learning (PBL) on critical thinking disposition in science education. *Journal of Environmental Protection and Ecology*, 14(3A), 1353–1361. https://www.researchgate.net/publication/286206344_Effect_of_problem-based_learning_PBL_on_critical_thinking_disposition_in_science_education.
- Facione, P. (1990a). The California critical thinking skills test—college level. Technical report# 1. Experimental validation and content validity. <https://files.eric.ed.gov/fulltext/ED327549.pdf>.
- Facione, P. (1990b). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction—The Delphi Report*. Millbrae, CA: California Academic Press. <https://philarchive.org/archive/FACCTA>.
- Facione, P., Facione, N., & Giancarlo, C. (2000a). *The California Critical Thinking Dispositions Inventory (CCTDI): the CCTDI Test Manual*. The California Academic Press.
- Facione, P., Facione, N., & Giancarlo, C. (2000b). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic*, 20(1), 61–84. <https://doi.org/10.22329/il.v20i1.2254>
- Facione, P. A., & Facione, N. C. (1992). *California critical thinking disposition inventory*. Millbrae, CA: California Academic Press.
- Gholami, M., Moghadam, P. K., Mohammadipoor, F., Tarahi, M. J., Sak, M., Toulabi, T., & Pour, A. H. H. (2016). Comparing the effects of problem-based learning and the traditional lecture method on critical thinking skills and metacognitive awareness in nursing students in a critical care nursing course. *Nurse Education Today*, 45, 16–21. <https://doi.org/10.1016/j.nedt.2016.06.007>
- Godzella, B. M., & Masten, W. G. (1998). Critical thinking and learning processes for students in two major fields. *Journal of Instructional Psychology*, 25(4), 256. <https://search.proquest.com/openview/68260a38af24aa0014505cf3880b214c/1?cbi=2029838&pq-origsite=gscholar>.
- Gomez, F. (2002). *Education as if people matter: A call for critical thinking and humanistic education* <https://www.open.uwi.edu/sites/default/files/bncdde/belize/conference/papers/gomez.html>.
- Halpern, D. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *The American psychologist*, 53, 449–455. <https://doi.org/10.1037/0003-066X.53.4.449>
- Halpern, D. (2001). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education*, 50, 270–286. <https://doi.org/10.1353/jge.2001.0024>
- Harbord, R. M., Harris, R. J., & Sterne, J. A. (2009). Updated tests for small-study effects in meta-analyses. *The Stata Journal*, 9(2), 197–210. <https://doi.org/10.1177/1536867X0900900202>
- He, Y., Gu, P., & Chen, L. (2009). PBL jiaoxuefa dui huli benkesheng pipanxingsiweingli yingxian de yanjiu [Study on influence of PBL teaching method on critical thinking ability of college nursing students]. *Chinese Nursing Research*, 23(4B), 1004–1005. <https://doi.org/10.3969/j.issn.1009-6493.2009.11.034>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.3102/10769986006002107>
- Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.
- Hesterberg, L. J. (2005). *Evaluation of a problem-based learning practice course: Do self-efficacy, critical thinking, and assessment skills improve?* (Publication Number 3162941) [Ph.D., University of Kentucky]. ProQuest One Academic. Ann Arbor. <https://search.proquest.com/dissertations-theses/evaluation-problem-based-learning-practice-course/docview/304992325/se-2?accountid=17096>.
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Howe, R.W., & Warren, C.R. (1989). *Teaching critical thinking through environmental education*. ERIC Clearinghouse. <https://eric.ed.gov/?id=ED324193>.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index? *Psychological methods*, 11(2), 193. <https://doi.org/10.1037/1082-989X.11.2.193>
- Hung, T. M., Tang, L. C., & Ko, C. J. (2015). How mental health nurses improve their critical thinking through problem-based learning. *Journal for Nurses in Professional Development*, 31(3), 170–175. <https://doi.org/10.1097/nnd.0000000000000167>
- Hursen, C. (2020). The effect of problem-based learning method supported by web 2.0 tools on academic achievement and critical thinking skills in teacher education. *Technology, Knowledge and Learning*, 26, 515–533. <https://doi.org/10.1007/s10758-020-09458-2>
- International Commission on the Futures of Education Commission. (2021). Reimagining our futures together: a new social contract for education. UNESCO. <https://unesdoc.unesco.org/notice?id=p:usmarcdef.0000379707>.
- Iwaoka, W., Li, Y., & Rhee, W. (2010). Research in food science education: Measuring gains in critical thinking in food science and human nutrition courses: The cornell critical thinking test, problem-based learning activities, and student journal entries. *Journal of Food Science Education*, 9, 68–75. <https://doi.org/10.1111/j.1541-4329.2010.01001.x>
- Jun, W. H., & Lee, E. (2013). Effects of S-PBL in fundamental nursing practicum among nursing students: comparison analysis of a ordinary least square and a quantile regression for critical thinking disposition. *The Journal of the Korea Contents Association*, 13(11), 1036–1045. <https://doi.org/10.5392/JKCA.2013.13.11.1036>
- Kim, D. H., Lee, Y., Hwang, M. S., Park, J. H., Kim, H. S., & Cha, H. G. (2012). Effects of a simulation-based integrated clinical practice program (SICPP) on the problem solving process, clinical competence and critical thinking in a nursing student. *The Journal of Korean Academic Society of Nursing Education*, 18(3), 499–509. <https://doi.org/10.5977/jkasne.2012.18.3.499>
- Kim, K. N. (2009). *Exploring undergraduate students' active learning for enhancing their critical thinking and learning in a large class* (Publication Number 3380934) *The Pennsylvania State University*. ProQuest Dissertations Publishing. <https://search.proquest.com/openview/f5d41d39e847d869e2f7597236c3f9c3/1?cbi=18750&diss=y&pq-origsite=gscholar>.
- Kong, L., Qin, B., Zhou, Y., Mou, S., & Gao, H. (2014). The effectiveness of problem-based learning on development of nursing students' critical thinking: A systematic review and meta-analysis. *International Journal of Nursing Studies*, 51(3), 458–469. <https://doi.org/10.1016/j.ijnurstu.2013.06.009>
- Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571350>
- Lave, J., & Wenger, E. (1991). Learning in doing: Social, cognitive, and computational perspectives. *Situated learning: Legitimate peripheral participation*, 10. <https://courses.cs.washington.edu/courses/cse590w/06au/resources/mcdermott-93.pdf>.
- Lawson, T. J. (1999). Assessing psychological critical thinking as a learning outcome for psychology majors. *Teaching of Psychology*, 26(3), 207–209. <https://doi.org/10.1207/s15328023top260311>
- Lee, J., Lee, Y., Gong, S., Bae, J., & Choi, M. (2016). A meta-analysis of the effects of non-traditional teaching methods on the critical thinking abilities of nursing students. *Bmc Medical Education*, 16(1), 1–9. <https://doi.org/10.1186/s12909-016-0761-7>
- Lee, J., & Son, H. K. (2021). Comparison of learning transfer using simulation problem-based learning and demonstration: An application of papanicolaou smear nursing education. *International Journal of Environmental Research and Public Health*, 18(4), 1765. <https://doi.org/10.3390/ijerph18041765>
- Leng, J., & Lu, X. (2020). Pipanxingsiwei zhende kejiao ma-jiyu 79 pian shiyan huo zhunshiyuan yanjiu de yuanfenxi [Can Critical Thinking Really Be Taught? A meta-analysis based on 79 experimental and quasi-experimental studies]. *Open Education Research*, 26(6), 110–118. <https://doi.org/10.13966/j.cnki.kfjyyj.2020.06.011>
- Li, J., & Pan, S. (2019). Dongxifang kexue kecheng biao zhun zhong pipanxingsiwei de bijiao fenxi [The comparison on critical thinking in science curricula of eastern and western Countries]. *Education Approach*, 1(1), 48–55. <http://www.cnki.com.cn/Article/CJFDTotal-SZTY201901007.htm>.
- Lin, C. F., Lu, M. S., Chung, C. C., & Yang, C. M. (2010). A comparison of problem-based learning and conventional teaching in nursing ethics education. *Nursing Ethics*, 17(3), 373–382. <https://doi.org/10.1177/0969733009355380>

- Liu, Z., Wu, W., & Jiang, Q. (2020). Wentishi xuexi dui daxuesheng pipanxingsiwei de yingxiang yanjiu-jiyu guoneiwai 31 xiang yanjiu de yuanfenxi [The Influence of PBL on Critical Thinking for Undergraduates: a meta-analysis based on 31 domestic and abroad studies]. *Higher Education Exploration*, (3), 43–49. <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=GJTA202003008&DbName=CJFQ2020>.
- Lyons, E. (2008). Examining the effects of problem-based learning and NCLEX-RN scores on the critical thinking skills of associate degree nursing students in a southeastern community college. *International journal of nursing education scholarship*, 5. <https://doi.org/10.2202/1548-923X.1524>. Article21.
- Mandeville, D., & Stoner, M. (2015). Research and teaching: Assessing the effect of problem-based learning on undergraduate student learning in biomechanics. *Journal of College Science Teaching*, 045. https://doi.org/10.2505/4/jcst15_045_01_66
- Mao, H., Ni, H., Yao, M., & Fu, B. (2017). ACCT lianhe PBL jiaoxuefa zai waike linchuang jiaoxue zhong dui husheng pipanxingsiweingli de yingxiang [Effect of ACCT combined with PBL teaching method on critical thinking ability of nursing students in surgical clinical teaching]. *Journal of Modern Medical Health*, 33(17), 2705–2707. <https://doi.org/10.3969/j.issn.1009-5519.2017.17.049>
- Martyn, J., Terwijn, R., Kek, M., & Huijser, H. (2014). Exploring the relationships between teaching, approaches to learning and critical thinking in a problem-based learning foundation nursing course. *Nurse Education Today*, 34(5), 829–835. <https://doi.org/10.1016/j.nedt.2013.04.023>
- Masek, A., & Yamin, S. (2011). The effect of problem based learning on critical thinking ability: A theoretical and empirical review. *International Review of Social Sciences and Humanities*, 2(1), 215–221. <https://www.researchgate.net/publication/266279314>.
- McMillan, J. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, 26, 3–29. <https://doi.org/10.1007/BF00991931>
- Muehlenkamp, J. J., Weiss, N., & Hansen, M. (2015). Problem-based learning for introductory psychology: Preliminary supporting evidence. *Scholarship of Teaching and Learning in Psychology*, 1(2), 125–136. <https://doi.org/10.1037/stl0000027>
- Nargundkar, S., Samaddar, S., & Mukhopadhyay, S. (2014). A guided problem-based learning (PBL) approach: Impact on critical thinking. *Decision Sciences Journal of Innovative Education*, 12. <https://doi.org/10.1111/dsji.12030>
- Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Journal of Educational Research Review*, 9, 114–128. <https://doi.org/10.1016/j.edurev.2012.12.002>
- Norris, S. P. (2003). The meaning of critical thinking test performance: The effects of abilities and dispositions on scores. In F. DJ (Ed.), *Critical thinking and reasoning: Current research, theory, and practice* (pp. 315–329). Hampton Press, Inc.
- OECD. (2018). The future of education and skills: Education 2030. [https://www.oecd.org/education/2030/E2030%20Position%20Paper%20\(05.04.2018\).pdf](https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf).
- Oliveira, L. B. d., Díaz, L. J. R., Carbogim, F. d. C., Rodrigues, A. R. B., & Püschel, V. A. d. A. (2016). Effectiveness of teaching strategies on the development of critical thinking in undergraduate nursing students: a meta-analysis. *Revista da Escola de Enfermagem da USP*, 50(2), 355–364. <https://doi.org/10.1590/S0080-623420160000200023>
- Ozturk, C., Muslu, G. K., & Dicle, A. (2008). A comparison of problem-based and traditional education on nursing students' critical thinking dispositions. *Nurse Education Today*, 28(5), 627–632. <https://doi.org/10.1016/j.nedt.2007.10.001>
- Pan, Y., Hu, C., Chen, P., Zhu, S., Zhang, Y., Guo, L., ..., & Pan, P. (2015). Shenzhangbingxue linchuang PBL jiaoxue dui xuesheng pipanxingsiweingli de yingxiang [The influence of clinical PBL teaching on critical thinking ability of students in nephrology]. *CJITWN*, 16(12), 1086–1088. <https://kns.cnki.net/kcms/detail/detail.aspx?filename=JXSB201512024&dbcode=CJFQ&dbname=CJFD2015&v=3LHgkHIP8hJU0LZXUTrqf5WjYkYMNQ2fkcJAT1ZEFnPiUyeVf5DRm7IBr0AZK4TPN>.
- Pardamean, B. (2012). Measuring change in critical thinking skills of dental students educated in a PBL curriculum. *Journal of Dental Education*, 76, 443–453. <https://doi.org/10.1002/j.0022-0337.2012.76.4.tb05276.x>
- Paul, R. W., & Binker, A. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world*. Rohnert Park: Center for Critical Thinking and Moral Critique, Sonoma State University.
- Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities—a dispositional theory of thinking. *Merrill-Palmer Quarterly—Journal of Developmental Psychology*, 39(1), 1–21. <http://www.jstor.org/stable/23087298>.
- Qin, H., Ge, W., & Sun, X. (2013). PBL jiaoxuefa zai shiyan jiaoxue zhong dui xuesheng pipanxingsiweingli de peiyang [Application of PBL teaching method in cultivation of critical thinking ability of students in experimental teaching]. *Experimental Technology and Management*, 30(6), 121–123. <https://doi.org/10.16791/j.cnki.sjg.2013.06.037>
- Quinn, S., Hogan, M., Dwyer, C., Finn, P., & Fogarty, E. (2020). Development and validation of the student-educator negotiated critical thinking dispositions scale (SENCITDS). *Thinking Skills and Creativity*, 38, Article 100710. <https://doi.org/10.1016/j.tsc.2020.100710>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Journal of Psychological Bulletin*, 86(3), 638. <https://doi.org/10.1037/0033-2909.86.3.638>
- Salomon, G. (1994). *To be or not to be (mindful)*. New Orleans, LA: American Educational Research Association Meetings.
- Sanderson, H. L. (2008). *Comparison of problem based learning and traditional lecture instruction on critical thinking, knowledge, and application of strength and conditioning* (Publication Number 3337476) [Ed.D., The University of North Carolina at Greensboro]. ProQuest One Academic. Ann Arbor. <https://search.proquest.com/dissertations-theses/comparison-problem-based-learning-traditional/docview/288256919/se-2?accountid=17096>.
- Savery, J. R. (2015). Overview of problem-based learning: Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*, 9(2), 5–15. <https://doi.org/10.7771/1541-5015.1002>
- Semerci, N. (2006). The effect of problem-based learning on the critical thinking of students in the intellectual and ethical development unit. *Social Behavior and Personality*, 34(9), 1127–1135. <https://doi.org/10.2224/sbp.2006.34.9.1127>
- Sendag, S., & Odabasi, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, 53(1), 132–141. <https://doi.org/10.1016/j.compedu.2009.01.008>
- Silviariza, W., Sumarmi, S., & Handoyo, B. (2021). Improving critical thinking skills of geography students with spatial problem based learning (SPBL). *International Journal of Instruction*, 14, 133–152. <https://doi.org/10.29333/iji.2021.1438a>
- Silviariza, W. Y., Sumarmi, & Handoyo, B. (2020). Using of spatial problem based learning (SPBL) model in geography education for developing critical thinking skills [Article]. *Journal for the Education of Gifted Young Scientists*, 8(3), 1045–1060. <https://doi.org/10.17478/JEGYS.737219>
- Son, H. K. (2020). Effects of S-PBL in maternity nursing clinical practicum on learning attitude, metacognition, and critical thinking in nursing students: A Quasi-experimental design. *International Journal of Environmental Research and Public Health*, 17(21). <https://doi.org/10.3390/ijerph17217866>. Article 7866.
- Song, Y. A. (2020). Effects of integrative simulation practice on nursing knowledge, critical thinking, problem-solving ability, and immersion in problem-based learning among nursing students. *kjwhn*, 26(1), 61–71. <https://doi.org/10.4069/kjwhn.2020.03.15.1>
- Sosu, E. M. (2013). The development and psychometric validation of a critical thinking disposition scale. *Thinking Skills and Creativity*, 9, 107–119. <https://doi.org/10.1016/j.tsc.2012.09.002>
- Suebnuakarn, S., & Haddawy, P. (2006). Modeling individual and collaborative problem-solving in medical problem-based learning. *User Modeling and User-Adapted Interaction*, 16(3), 211–248. <https://doi.org/10.1007/s11257-006-9011-8>
- Sulaiman, F. (2011). *The effectiveness of Problem-Based Learning (PBL) online on students' creative and critical thinking in physics at tertiary level in Malaysia* [Doctoral, University of Waikato]. Hamilton, New Zealand. <https://hdl.handle.net/10289/4963>.
- Sun, B., Du, X., Emmersen, J., & Toft, E. (2013). PBL and critical thinking disposition in Chinese medical students -A randomized cross-sectional study. *Journal of Problem Based Learning in Higher Education*, 1, 2013–2072. <https://doi.org/10.5278/ojs.jpblhe.v1i1.275>
- Tayyeb, R. (2013). Effectiveness of problem based learning as an instructional tool for acquisition of content knowledge and promotion of critical thinking among medical students. *Journal of the College of Physicians and Surgeons-Pakistan: JCPSP*, 23, 42–46. <https://pubmed.ncbi.nlm.nih.gov/23286622/>.
- Temel, S. (2014). The effects of problem-based learning on pre-service teachers' critical thinking dispositions and perceptions of problem-solving ability. *South African Journal of Education*, 34(1), 1–20. <https://doi.org/10.15700/201412120936>. Article 769.
- Thompson, C., & Rebesch, L. M. (1999). Critical thinking skills of baccalaureate nursing students at program entry and exit. *Nursing and Health Care Perspectives*, 20(5), 248–252. <https://pubmed.ncbi.nlm.nih.gov/10754847/>.

- Tian, S., Wang, L., & Qiu, Y. (2018). Wentishi jiaoxuefa dui gongke daxuesheng pipanxingsiweiqingxiang yingxiang de shizheng yanjiu [An empirical study on the influence of problem-based learning on critical thinking disposition of engineering students]. *Research in Higher Education of Engineering*, (6), 156–160. <http://www.cnki.com.cn/Article/CJFDTotal-GDGJ201806028.htm>.
- Tiwari, A., Lai, P., So, M., & Yuen, K. (2006). A comparison of the effects of problem-based learning and lecturing on the development of students' critical thinking. *Medical Education*, 40(6), 547–554. <https://doi.org/10.1111/j.1365-2929.2006.02481.x>
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times* (1 ed.). Jossey-Bass.
- Tseng, H.-C., Chou, F.-H., Wang, H.-H., Ko, H.-K., Jian, S.-Y., & Weng, W.-C. (2011). The effectiveness of problem-based learning and concept mapping among Taiwanese registered nursing students. *Nurse Education Today*, 31(8), e41–e46. <https://doi.org/10.1016/j.nedt.2010.11.020>
- Ulger, K. (2018). The effect of problem-based learning on the creative thinking and critical thinking disposition of students in visual arts education. *Interdisciplinary Journal of Problem-Based Learning*, 12. <https://doi.org/10.7771/1541-5015.1649>
- Velde, B. P., Wittman, P. P., & Vos, P. (2006). Development of critical thinking in occupational therapy students. *Occupational Therapy International*, 13(1), 49–60. <https://doi.org/10.1002/oti.20>
- Wang, H., Zheng, X., & Meng, Y. (2009). Effect of problem-based learning in nursing students: A meta-analysis. *Chinese Journal of Evidence-Based Medicine*, 09(1), 93–98. <https://doi.org/10.7507/1672-2531.20090019>
- Wang, M. (2009). *Yi wenti wei jichu de xuexi zai jichu hulixue jiaoxue zhong de yingyong* [Dissertation, Qingdao University]. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMPD&dbname=CMPD2009&filename=2009128459.nh&v=Dzbbtx8rKZUO60XYDsO8NvBWomaif1rXC5AqP6LZbOYigpJH2gLMrfS3157Kgti>.
- Wang, M., & Han, Y. (2019). Wentishi xuexi jiaoxue fangfa zai jianyanke linchuang daijiao de xiaoguo [Effect of problem based learning teaching method in clinical teaching of laboratory]. *Journal of Preventive Medicine Chi PLA*, 37(10), 184–185. <https://doi.org/10.13704/j.cnki.jyyx.2019.10.068>
- Wang, Y. (2018). The influence of PBL teaching mode on critical thinking ability for non-english majors under network environment. *MATEC Web of Conferences*, 228, 05022. <https://doi.org/10.1051/mateconf/201822805022>
- Watson, G. (1980). *Watson-Glaser critical thinking appraisal*. San Antonio, TX: Psychological Corporation.
- Weissinger, P. (2003). *Critical thinking skills of first-year dental students enrolled in a hybrid curriculum with a problem-based learning component* [Dissertation, Indiana University]. United States. <https://www.proquest.com/openview/67905566d61505952b0b90c6d87c8109/1?cbl=18750&diss=y&pq-origsite=gscholar>.
- Wiles, J., & Bondi, J. (1989). *Curriculum development: A guide to practice* (3 ed.). Macmillan Publishing Company.
- Willingham, D. (2010). Critical thinking why is it so hard to teach? *Arts Education Policy Review*, 109. <https://doi.org/10.3200/AEPR.109.4.21-32>
- Yoo, M. R., Choi, Y. J., & Kang, M. S. (2009). The effects of PBL (Problem-Based learning) on the self-directed learning, critical thinking disposition, and problem solving process of nursing students. *Journal of Korean Acadamey Fundam Nursing*, 16(1). <https://www.koreascience.or.kr/article/JAKO200919463950845.page>.
- Yoon, J. (2004). *Development of an instrument for the measurement of critical thinking disposition in nursing* [Dissertation, The Catholic University of Education]. Seoul, Korea.
- Yu, D., Zhang, Y., Xu, Y., Wu, J., & Wang, C. (2013). Improvement in critical thinking dispositions of undergraduate nursing students through problem-based learning: A crossover-experimental study. *Journal of Nursing Education*, 52(10), 574–581. <https://doi.org/10.3928/01484834-20130924-02>
- Yuan, H. (2007). A meta-analysis of the effect of problem based learning on critical thinking. *Chinese Nursing Research*, 13. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2007&filename= SXHZ200713046&v=kOHZBtQ7IYbjSRwvsZgkdyt0jHgEyki34%25mmd2FuQFn9uPFQoFoaNOTY1zshg5evQVJAX>.
- 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, L. (2014). *Wentishi xuexi dui yingyu zhuanye xuesheng pipanxingsiwei de yingxiang yanjiu* [A Study of the Effect of Problem-based Learning on English Majors' Critical Thinking] [Dissertation, Shenyang Normal University]. CNKI. Shenyang, China. <http://cdmd.cnki.com.cn/Article/CDMD-10166-1014231387.htm>.
- Zhao, W., & Liu, X. (2011). Jiyu wangluo de PBL moshi zai laonian hulixue jiaoxue zhong de yingyong [Application of web based PBL teaching for the course of Gerontological Nursing]. *Journal of Nursing Science*, 26(15), 64–66. <https://doi.org/10.3870/hlxxz.2011.15.064>
- Zhu, X. (2010). *Gaozhi gaozhuan huli xuesheng pingpanxing siweinengli xiankuang diaocha ji peiyang duice yanjiu* [Critical thinking ability investigation and teaching countermeasures research of nursing students from higher vocational schools and colleges] [Dissertation, Zhejiang University]. CNKI. Hangzhou, China. <http://cdmd.cnki.com.cn/article/cdmd-10335-1011057681.htm>.