REVIEW ARTICLE

Situating Higher-Order, Critical, and Critical-Analytic Thinking in Problem- and Project-Based Learning Environments: A Systematic Review

Sofie M. M. Loyens[1] ⓘ · **Julianne E. van Meerten**[2] · **Lydia Schaap**[3] · **Lisette Wijnia**[4]

Abstract

Critical thinking (CT) is widely regarded as an important competence to obtain in education. Students' exposure to problems and collaboration have been proven helpful in promoting CT processes. These elements are present in student-centered instructional environments such as problem-based and project-based learning (P(j) BL). Next to CT, also higher-order thinking (HOT) and critical-analytic thinking (CAT) contain elements that are present in and fostered by P(j)BL. However, HOT, CT, and CAT definitions are often ill-defined and overlap. The present systematic review, therefore, investigated how HOT, CT, and CAT were conceptualized in P(j) BL environments. Another aim of this study was to review the evidence on the effectiveness of P(j)BL environments in fostering HOT, CT, or CAT. Results demonstrated an absence of CAT in P(j)BL research and a stronger focus on CT processes than CT dispositions (i.e., trait-like tendency or willingness to engage in CT). Further, while we found positive effects of P(j)BL on HOT and CT, there was a lack of clarity and consistency in how researchers conceptualized and measured these forms of thinking. Also, essential components of P(j)BL were often overlooked. Finally, we identified various design issues in effect studies, such as the lack of control groups, that bring the reported outcomes of those investigations into question.

Keywords Problem-based learning · Project-based learning · Higher-order thinking · Critical thinking · Critical-analytic thinking

✉ Sofie M. M. Loyens
s.loyens@ucr.nl

1   University College Roosevelt, Utrecht University, Utrecht, The Netherlands

2   Department of Human Development & Quantitative Methodology, College of Education, University of Maryland, College Park, MD, USA

3   HU University of Applied Sciences, Utrecht, The Netherlands

4   Faculty of Educational Sciences, Open Universiteit, Heerlen, The Netherlands

Critical thinking (CT) is widely regarded as an important competence to learn, and its importance has only increased over time (Pellegrino & Hilton, 2012). Mastery of this ability is not only necessary for students but also for working professionals and informed citizens (Bezanilla et al., 2021). Therefore, thinking critically is a central aim of education (e.g., Butler & Halpern, 2020). A meta-analysis demonstrated that the opportunity for dialogue, exposing students to authentic or situated problems and examples, and mentoring them in these habits of mind positively affected CT (Abrami et al., 2015). All of these pedagogical elements are, to some extent, present in student-centered instructional environments. Thus, it would seem that a good way to teach CT is by using active, student-centered instructional methods (Bezanilla et al., 2021; Lombardi et al., 2021). Problem-based learning (PBL) and project-based learning (PjBL) are prototypical examples of active, student-centered instructional methods. Next to CT, higher-order thinking (HOT) and critical-analytic thinking (CAT) also contain elements that are present in and fostered by student-centered learning environments, such as PBL and PjBL. However, CT, HOT, and CAT definitions are often ill-defined and overlap.

For the aforementioned reasons, the present study investigated CT, HOT, and CAT in student-centered learning environments. We will focus on PBL and PjBL because these two formats are most frequently studied in the research literature and hence constitute the vast majority of student-centered instructional methods (Authors, 2022; Nagarajan & Overton, 2019). In addition, both PBL and PjBL have been included as acknowledged instructional formats in the Cambridge Handbook of Learning Sciences (Sawyer, 2014) and have specific criteria that need to be fulfilled to be labeled as PBL and PjBL. Despite their unique characteristics and origin, PBL and PjBL share common ground because of their joint roots in constructivist learning theory (Authors, 2022; Loyens & Rikers, 2017). PBL and PjBL as pedagogies can be seen as manifestations of constructivist learning. Constructivism is a theory or view on how learning happens, which holds that learners construct knowledge out of experiences. It has roots in philosophy and stresses the student's active role in his or her knowledge-acquisition process (Loyens et al., 2012).

What is PBL?

There are many definitions of PBL in the research literature, all presenting different perspectives and ideas with regard to this educational pedagogy (Dolmans et al., 2016; Zabit, 2010). Notwithstanding the variety of definitions, different PBL implementations demonstrate some shared characteristics. Based on the original method developed at McMaster University (Spaulding, 1969), Barrows (1996) described six core characteristics of PBL. The first characteristic is that learning is student centered. Second, learning occurs in small student groups under the guidance of a tutor. The third characteristic refers to the tutor as a facilitator or guide. Fourth, students encounter so-called "authentic" (i.e., relating to real life) problems in the learning sequence before any preparation or study has occurred. Fifth, these problems function as triggers for students' prior knowledge activations, which leads to the discovery of knowledge gaps. Finally, students overcome these knowledge gaps through

self-directed learning, which requires sufficient time for self-study (Schmidt et al., 2009).

Besides the core elements, different PBL environments share similarities in terms of the process. For example, Wijnia et al. (2019) highlighted that PBL as a process consists of three separate stages: an initial discussion phase, a self-study phase, and a reporting stage. First, students are given a meaningful problem that describes an observable phenomenon or event. The instructional goal of the problem presented to the students can differ. For example, the problem could originate from professional practice or be related directly to distinctive events in a particular domain or field of study. An example of a problem related to a specific domain of study from an introductory psychology course reads as follows (Schmidt et al., 2007):

Coming home from work, tired and in need of a hot bath, Anita, an account manager, discovers two spiders in her tub. She shrinks back, screams, and runs away. Her heart pounds, a cold sweat is coming over her. A neighbor saves her from her difficult situation by killing the little animals using a newspaper." Explain what has happened here. (p. 92)

During the first stage, prior knowledge is important as students come up with theories to explain the problem based on their life experiences. Because their knowledge is often limited and insufficient, students formulate learning issues (formulated as questions) to guide their research and further self-study. All this takes place in a class discussion, usually in classes with fewer than 12 students. During the second stage, students consult learning resources to gain knowledge relevant to the problem and to address the learning issue questions. These resources can be selected by the students, the tutor, or a combination of both (Wijnia et al., 2019). In tandem with these steps, students have to plan and monitor study activities that need to be carried out before the next class meeting (Loyens et al., 2008). In the final stage, students reconvene under the guidance of their tutor to share and evaluate their findings critically and elaborate on their newly acquired knowledge. Students apply this knowledge to the problem to identify plausible solutions or explanations (Loyens & Rikers, 2017; Wijnia et al., 2019).

What is PjBL?

In project-based or project-centered learning (PjBL), the learning process is organized around activities that drive students' actions (Blumenfeld et al., 1991). Students learn central concepts and principles of a discipline through the projects. This "learning-by-doing" approach of PjBL could help motivate students to learn, as they play an active role in the process (Saad & Zainudin, 2022). Students have a significant degree of control over the project they will work on and what they will do in the project. The projects are hence student-driven and, similar to PBL, are intended to generate learner agency. Specific end products need to result from the work, although the processes to get to the end product can vary. The end products (e.g., a website, presentation, or report) serve as the basis for discussion, feedback, and revision (Blumenfeld et al., 1991; Helle et al., 2006; Tal et al., 2006). Also, even

though different forms of PjBL exist, most start with a driving question or problem and typically incorporate the following features (Krajcik, 2015; Thomas, 2000):

1. Projects for which the students seek solutions or clarifications are relevant to their lives.
2. PjBL involves planning and performing investigations to answer questions.
3. Students collaborate with other students, teachers, and members of society.
4. PjBL is centered around producing artifacts.
5. Technology is used when appropriate.

In sum, students perform a series of collaborative inquiry activities that should help them acquire new, domain-specific knowledge and thinking processes to solve real-world problems. The PjBL end products must reflect learners' knowledge of the project topic and their metacognitive knowledge (Grant & Branch, 2005). A project can be a problem to solve (e.g., How can we reduce the pollution in the schoolyard pond?), a phenomenon to investigate (e.g., Why do you stay on your skateboard?), a model to design (e.g., Create a scale model of an ideal high school), or a decision to make (e.g., Should the school board vote to build a new school?; Yetkiner et al., 2008).

Students work together and projects last for considerable periods (Helle et al., 2006). The role of the instructor consists of facilitating the project. That is, the instructor helps with framing and structuring the projects, monitors the development of the end product, and assesses what students have learned (Chiu, 2020; David, 2008; Helle et al., 2006).

What PBL and PjBL Foster

One aim of schools and colleges implementing student-centered approaches such as PBL and PjBL is to increase students' competence in tackling complex problems common in an ever-changing world (Gijbels et al., 2005). To that end, several goals and desired outcomes have been put forward for PBL and PjBL. The primary goal is to educate the students to a level where they can comfortably use and retrieve information when needed and identify situations where specific knowledge and strategic processes are applicable. With these strategies and knowledge, students can start developing plausible explanations of phenomena that represent important disciplinary understandings (Loyens et al., 2012; McNeill & Krajcik, 2011). Several studies have investigated the effectiveness of P(j)BL on knowledge acquisition (e.g., Chen & Yang, 2019; Strobel & Van Barneveld, 2009). In addition, both PBL and PjBL consist of collaborative learning sessions that could foster effective interpersonal communication. Such abilities can enable learners to contribute to discussions in clear and appropriate ways, help to reach conclusions and answers more easily, and identify inconsistencies and unresolved issues (Loyens et al., 2008, 2012).

Further, students could develop problem-solving strategies while working on problems or projects (Krajcik et al., 2008). Even when problems are highly complex

and ill-structured, they can be effectively analyzed, and plausible responses can be identified (Loyens et al., 2012). Also, because the problems and projects are specific to the students' domain of study, the knowledge and strategies they acquire are applicable to their future professional practice. Therefore, problems and projects are believed to be more engaging, motivating, and interesting for the students (Hmelo-Silver, 2004; Larmer et al., 2015; Saad & Zainudin, 2022). Finally, as noted, student-centered instructional methods such as PBL and PjBL imply a different, less directive role for the teacher. Consequently, students receive more during the learning process. The success of PBL and PjBL also rests on the "preparedness of a student to engage in learning activities defined by him- or herself, rather than by a teacher" (Schmidt, 2000, p. 243), a process referred to as self-directed learning.

Even though the research literature on P(j)BL does not explicitly state that these instructional formats should foster HOT, CT, and CAT, their design and implementations do appear to require students' engagement in these forms of thinking. Thus, we will seek to define these constructs within the context of student-centered learning environments.

HOT, CT, and CAT

Like PBL and PjBL, many definitions of HOT, CT, and CAT can be found in the literature. We acknowledge our inability to be complete regarding the different domains and traditions (i.e., philosophical, psychological, educational) of HOT, CT, and CAT, in which definitions have been put forward. Rather, we will focus on definitions that help describe the role of HOT, CT, and CAT in student-centered learning environments.

What Is Higher Order Thinking (HOT)?

First, HOT can be seen as an overarching concept defined as "skills that enhance the construction of deeper, conceptually-driven understanding" (Schraw & Robinson, 2011, p. 2). Framed in more traditional terms, HOT corresponds with Bloom's taxonomy, with remembering or recalling facts reflecting lower-order cognitive thinking (i.e., concerned with the acquisition of knowledge or information) and comprehending, applying, analyzing, synthesizing, and evaluating as higher-order thinking, referring to more intellectual abilities and skills (Lombardi, 2022; Miri et al., 2007). The focus on thinking skills does not imply that the essential importance of knowledge is abandoned. In fact, knowledge is needed and related to thinking processes, which comes to the fore in, for example, later revisions of Bloom's taxonomy (Lombardi, 2022).

HOT has been put forward as having four components (Schraw et al., 2011): (a) reasoning (i.e., induction and deduction), (b) argumentation (i.e., generating and evaluating evidence and arguments), (c) metacognition (i.e., thinking about and regulating one's thinking), and (d) problem solving and critical thinking (CT). Problem-solving involves several steps carried out consecutively: 1) identifying

and representing the problem at hand, 2) selecting and applying a suitable solution strategy, and 3) evaluating the process and solution (Chakravorty et al., 2008). CT refers to the reflective thinking that leads to certain outcomes (i.e., decision-making) and actions (Ennis, 1987). CT is, in this view, considered a subcomponent of HOT (Schraw et al., 2011). Indeed, in its measurement, CT has also addressed HOT processes such as analysis and synthesis (Lombardi, 2022).

Yen and Halili (2015) also characterize HOT as an umbrella term for all manner of reflective thinking, including creative thinking, problem-solving, decision-making, and metacognitive processing. Also, in this definition, several components refer to the taxonomy of Schraw and colleagues (2011): reflection (component d) and metacognition (component c), problem-solving (component d), and decision-making (CT, component d). The only exception is creative thinking, which is not included as a subcomponent by Schraw and colleagues (2011); however, they acknowledge that this could be part of a broader taxonomy of HOT, together with, for example, moral reasoning.

What Is CT?

The literature on CT traces back to the Greek philosophers who sought to explain the origin and meaning of such thinking. As Van Peppen (2020) points out, "the word critical derives from the Greek words 'kritikos' (i.e., to judge/discern) and 'kriterion' (i.e., standards)," and hence "CT implies making judgments based on standards" (p. 11). A second important ancestor of CT was John Dewey, who spoke of "reflective thinking" when referring to CT. From thereon, many traditions and definitions of CT have been formulated. Ennis (1987), for example, defined CT as the thinking process focused on the decision of what to believe or what to do. He further expanded the idea of Glaser (1941), who acknowledged the role of dispositions in CT. Ennis (1962) distinguished two distinct CT components, dispositions and abilities, with the first one being more trait-like tendencies (e.g., dispositions toward inquisitiveness, open-mindedness, sensitivity to other points of view, cognitive flexibility) and the second referring to actual cognitive activities (e.g., focusing, analyzing arguments, asking questions, evaluating evidence, comparing potential outcomes; Schraw et al., 2011).

Scholars of the American Philosophical Association tried to come up with a consensus on the definition of CT, referred to as the Delphi Panel (Facione, 1990b). The processes associated with CT in that report were interpretation (i.e., understanding and articulating meaning), analysis (i.e., identifying relationships between information, including argument analysis), evaluation (i.e., making judgments and assessments about the credibility of information, including assessing arguments), inference (i.e., identifying the necessary information for decision making, including coming up with hypotheses), explanation (i.e., articulating and presenting one's position, arguments, and analysis used to determine that position), and self-regulation (i.e., self-analyzing and examining one's inferences and correcting when necessary; Facione, 1990b). The latter component (i.e., self-regulation) has a strong metacognitive character (Zimmerman & Moylan, 2009).

Finally, another conceptualization of CT that resulted in the development of a widely used measurement instrument for CT comes from Halpern. Halpern (2014) mainly focuses on abilities/cognitive activities and less on dispositions in her definition of CT. Specifically, she states that CT entails "cognitive skills or strategies that increase the probability of a desirable outcome" (p. 8). Halpern's taxonomy consists of five main elements; verbal reasoning, argument analysis, hypothesis testing, likelihood and uncertainty, and decision-making/problem-solving.

What Is CAT?

CAT refers to the processes we use "when we question or at least do not simply passively accept the accuracy of claims as givens" (Byrnes & Dunbar, 2014, p. 479). Its distinguishing feature compared to CT is its focus on justification and determining whether appropriate and credible evidence supports a claim or proposed response (Murphy et al., 2014). Although some frameworks of HOT and CT include the component of analysis and evaluation, the CAT research literature puts the processes of "weighing the evidence" at the forefront. Alexander (2014) describes the process of CAT in four consecutive steps. The process starts with a claim or task for which one collects data or evidence. Individuals then evaluate or judge these data or evidence and, as the last step, integrate it with their knowledge and beliefs. Dispositions are not considered in the CAT literature, although individual differences, such as prior knowledge and goals, can act as moderators.

In sum, unequivocal definitions of HOT, CT, and CAT are hard to find, although definitions do share important attributes. As Byrnes and Dunbar (2014) point out, "operational definitions follow from theoretical definitions" (p. 482). Indeed, the definitions introduced in this overview have led to several operationalizations and measurements of HOT, CT, and CAT. Those operationalizations are also important in the discussion on how HOT, CT, and CAT are framed within student-centered learning and whether P(j)BL might foster such valued forms of thinking.

The Link Between HOT, CT, and CAT and P(j)BL

To establish the link between HOT, CT, and CAT and the learning environments P(j)BL, we carefully examined two lines of research literature: literature on how to effectively teach HOT, CT, and CAT, as well as literature on the learning processes involved in P(j)BL (i.e., a synthesis between cognitive and instructional science). Most of the research directed toward teaching reflective forms of thinking have addressed CT as a form of HOT (e.g., Abrami et al., 2015; Miri et al., 2007; Schraw et al., 2011). For example, Miri and colleagues (2007) defined three teaching strategies that should encourage students to engage collaboratively in CT-aligned processes (e.g., asking appropriate questions and seeking plausible solutions). Those teaching strategies are (a) dealing with real cases in class, (b) encouraging class discussions, and (c) fostering inquiry-oriented experiments.

The link with P(j)BL is evident, as they are centered around dealing with real problems and cases in collaborative class discussions. Certainly, PjBL sometimes requires the execution of experiments. The sharing of knowledge and collaboration has a place in both instructional formats, in PBL in the reporting phase, in PjBL because students collaborate with other students, teachers, and members of society. In their meta-analysis, Abrami and colleagues (2015) reviewed possible instructional strategies that could foster CT. They concluded that two types of interventions helped develop CT processes: discussion and "authentic or situated problems and examples … particularly when applied problem solving … is used" (Abrami et al., 2015, p. 302). Again, there is a clear link because using authentic problems and class discussions are central components of PBL and PjBL. Also, Torff (2011) labeled core P(j)BL activities as "high-CT activities" (p. 363): Socratic discussion, debate, problem-solving, problem finding, brainstorming, decision-making, and analysis.

With respect to CAT, Byrnes and Dunbar (2014) put forward some instructional approaches that should prove facilitative. In their view, "students should pose unanswered questions that require the collection of data or evidence" (p. 488). Subsequently, they need to "engage in appropriate methodologies to gather this evidence." Finally, students need to "have opportunities to be surprised by unanticipated findings and discuss or debate how the anticipated, unanticipated, and missing evidence should be interpreted" (p. 488). These authors also stress the importance of working in teams, engaging in discussions, identifying sources of uncertainty and problems of interpretation, and presenting findings and conclusions for peer review. Also, regarding CAT, parallels can be drawn with PBL and PjBL in which a problem or question is the starting point, and students engage in several learning activities to develop an understanding and potentially a solution.

The second line of research that is useful to establish the link between HOT, CT, and CAT and P(j)BL, deals withs learning processes. For example, Krajcik et al. (2008) argued that processes common to project-based approaches involve learners in "scientific practices such as argumentation, explanation, scientific modeling, and engineering design" (p. 3). Furthermore, Krajcik and colleagues mention that students learning in these environments use problem-solving, design, decision-making, argumentation, weighing of different pieces of evidence, explanation, investigation, and modeling. Some scholars mention the development of metacognitive knowledge as an outcome of PjBL (Grant & Branch, 2005).

A similar case can be made for PBL. Hung, W. and colleagues (2008) indicated that students who experience PBL possess better hypothesis-testing abilities due to their more coherent explanations of hypotheses and hypothesis-driven reasoning. Further, the PBL process relies heavily on group discussions of real-life problems, discovering knowledge gaps, gathering information/evidence to answer the learning issues/questions, analyzing the evidence, resolving unclarities, and deciding on the outcome.

The Present Study

The present study aimed to situate HOT, CT, and CAT in PBL and PjBL environments. As we have set out to establish, even though fostering HOT, CT, and CAT may not be an explicit goal of these student-centered approaches; there are theoretical and empirical reasons to expect an association to exist. The research literature on how to effectively teach HOT, CT, and CAT has mentioned instructional formats that use discussion and problem-solving. Secondly, the research literature on the learning processes that take place in student-centered learning environments like P(j)BL mention HOT, CT, and CAT processes such as decision making, argumentation, weighing of different pieces of evidence, explanation, investigation (Abrami et al., 2015; Krajcik et al., 2008). Therefore, the first research question we posed was: *How are HOT, CT, and CAT conceptualized in student-centered learning environments?* In addition, parallels exist between processes involved in HOT, CT, and CAT on the one hand and the learning activities/processes in P(j)BL on the other hand. Moreover, effective instructional activities to foster HOT, CT, and CAT (e.g., Abrami et al., 2015) are core activities in P(j)BL. Therefore, the second aim of this study was *to review the evidence on the effectiveness of student-centered environments in fostering either HOT, CT,* or *CAT.* To that end, we carried out a review study of studies investigating HOT, CT, and CAT in the context of PBL and PjBL.

Method

Search Strategy

For this review, we systematically investigated six online databases: Web of Science (Core Collection) and five EBSCO databases (ERIC, Medline, PsycInfo, Psychology and Behavioral Sciences, and Teacher Reference Center). We included Medline in this list because problem-based learning originated in medical education (Spaulding, 1969, see also Servant-Miklos, 2019) and is often researched in the context of medical education (W. Hung et al., 2019; Koh et al., 2008; Smits et al., 2002; Strobel & Van Barneveld, 2009). For this systematic review, we used the following Boolean string of search terms (Oliver, 2012): "project based learning" or "project based instruction" or "project based approach" or "PjBL" OR "problem based learning" or "problem based approach" or "problem based instruction" or "PBL" AND "higher order thinking" or "critical thinking" or "critical analytic* thinking."[1] Because of its medical connotations (i.e., it is also an abbreviation for "peripheral blood lymphocytes"; e.g., Caldwell et al., 1998), "PBL" was not used for our Web of Science search. The search terms match our research questions in that they include both PBL and PjBL, as well as HOT, CT, and CAT, thus leading us to those studies

---

[1] Note that in the search terms for the student-centered learning environments, we did not use a hyphen, that is, "problem based learning" instead of "problem-based learning".

that included those variables. In addition to using the aforementioned terms, we delimited our search to peer-reviewed records written in English.

Selection Process

Inclusion Criteria

In addition to the search parameters established by our search terms, we used the following inclusion criteria to determine our final sample—studies had to: (a) use a quantitative measure of HOT, CT, or CAT, (b) take place in a PBL or PjBL environment, (c) be an empirical study that took place in a classroom context, (d) investigate K-12 or higher education students, and (e) be published as a peer-reviewed journal article. For example, to be included in our review, studies must have a dedicated measure of HOT, CT, or CAT that gives insight into the authors' conceptualization of these constructs. The measure used could be a standardized instrument or one that was researcher-designed.

Also, in line with our research aims, we only included studies that focused on PBL or PjBL and met the basic criteria for these student-centered approaches. To judge the quality of this learning environment, the definition or description in the theoretical framework and the implementation of those environments were assessed against the defining characteristics established in the literature (Barrows, 1996; Hmelo-Silver, 2004; Schmidt et al., 2009). For PBL, those criteria included (a) student-centered, active learning, (b) the guiding role of teachers, (c) collaborative learning in small groups, (d) the use of realistic problems as the start of the learning process, and (e) ample time for (self-directed) self-study. Furthermore, PBL had to contain the three process phases (i.e., initial discussion, self-study, and reporting phases). For PjBL, those criteria included (a) the project starts with a driving question or problem, (b) the project is relevant and authentic, (c) collaborative, inquiry learning activities take place, (d) room for student autonomy and the guiding role of teachers, (e) the project is central to the curriculum, and (f) the creation of a tangible product (Authors, 2022). We only included studies in a classroom context in K-12 and higher education because we wanted to focus on the effectiveness of these learning environments in formal educational settings. We focused on peer-review articles to better ensure the quality of the study.

Exclusion Criteria

We excluded self-report items (often from course evaluations) such as, "I have improved my ability to judge the value of new information or evidence presented to me" or "I have learned more about how to justify why certain procedures are undertaken in my subject area" (Castle, 2006). We also excluded the critical thinking scale from the Motivated Strategies for Learning Questionnaire (MSLQ; Duncan & McKeachie, 2005). In the MSLQ, critical thinking is viewed as a learning strategy. In light of strong criticism toward the reliance on MSLQ to effectively gauge these forms of

reflective thinking (see Dinsmore & Fryer, 2022, this issue), we have excluded these studies (e.g., Sungur & Tekkaya, 2006).

We also excluded studies that did not meet the defining criteria for PBL or PjBL. For example, if a study indicated investigating PBL where the learning process started with a lecture instead of a problem. Furthermore, we excluded studies that did not adequately describe the process the researchers labeled as PBL or PjBL (e.g., Razali et al., 2017). We also excluded studies that consisted of a combination of P(j)BL with additional activities (e.g., concept maps with PBL; Si et al., 2019) or interventions (e.g., a CT or motivation intervention combined with PBL; Olivares et al., 2013).

Moreover, we excluded studies that took place in a laboratory setting, intervention studies at places such as summer camps, tutoring, afterschool programs or studies with employee samples (e.g., health nurses; T.-M. Hung et al., 2015), because of our focus on formal education. We further excluded theoretical, conceptual, or "best practices" articles. Finally, for this review, we excluded peer-reviewed conference papers and abstracts ("wrong format") as they were often hard to retrieve or provided too little information to code the outcome measures and learning environment.

Coding and Final Sample

Figure 1 outlines our entire search process, in line with PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines (Moher et al., 2015). As can be seen there, our initial searches in the Web of Science and EBSCO databases provided us with 2,968 results, which we uploaded to the Rayyan platform (Ouzzani et al., 2016). After removing duplicates identified by the Rayyan platform, 2,545 papers remained, of which we screened the titles, abstracts, and, if needed, the full texts. We identified an additional 27 duplicates and excluded 2,405 papers because the studies did not meet our inclusion criteria.

We selected 113 studies for further inspection and coding. The specific codes were in line with our inclusion and exclusion criteria and were designed to streamline the final elimination round (i.e., excluding articles after in-depth reading). Table 1 provides a detailed overview of the codes used.

The search process overall resulted in a final sample of 28 studies. Of these 113 studies, 84 studies were excluded (see Fig. 1 for exclusion reasons). We excluded one paper because it contained duplicate data (Yuan et al., 2008b) reported in another paper (Yuan et al., 2008a). We selected the Yuan et al. (2008a) paper instead of the Yuan et al. (2008b) paper because the former also reported the results of the control group (i.e., lecture-based learning). In contrast, the latter only included the data of the PBL group.

Results and Discussion

Descriptives of the Final Sample

Before answering our research questions, we describe the characteristics of the 28 included studies (see Table 2). Of the included studies, 22 investigated a higher
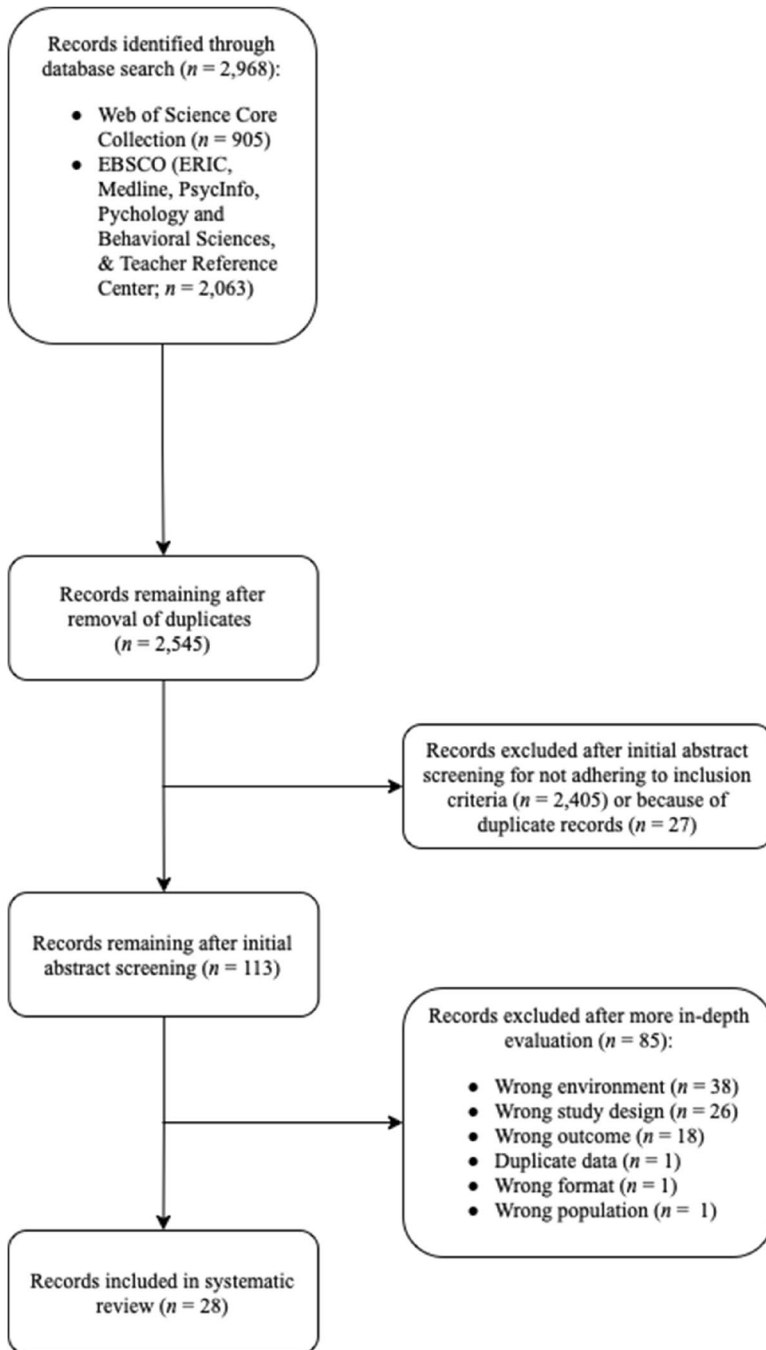
Records identified through
database search (*n* = 2,968):

- Web of Science Core
  Collection (*n* = 905)
- EBSCO (ERIC,
  Medline, PsycInfo,
  Pychology and
  Behavioral Sciences,
  & Teacher Reference
  Center; *n* = 2,063)

Records remaining after
removal of duplicates
(*n* = 2,545)

Records excluded after initial abstract
screening for not adhering to inclusion
criteria (*n* = 2,405) or because of
duplicate records (*n* = 27)

Records remaining after initial
abstract screening (*n* = 113)

Records excluded after more in-depth
evaluation (*n* = 85):

- Wrong environment (*n* = 38)
- Wrong study design (*n* = 26)
- Wrong outcome (*n* = 18)
- Duplicate data (*n* = 1)
- Wrong format (*n* = 1)
- Wrong population (*n* = 1)

Records included in systematic
review (*n* = 28)

Fig. 1 Flowchart of the Search and Selection Process

Table 1 Coding Scheme

Category	Specifics
Higher-order, critical, or critical-analytic thinking	Label (HOT/CT/CAT) Definition or conceptualization Reference
Learning environment	Label (PBL/PjBL) Definition or conceptualization Implementation A score based on adherence to established quality criteria (1 = exclude, 2 = maybe, 3 = include)
Academic domain	STEAM/Health Sciences/Other
Research question or aim	The research question/aim reported in the paper
Research design	e.g., independents group pre-post, pre-post, independent groups, longitudinal design
Control group	Yes/no Descriptives
Duration of the intervention	Number of hours per week/number of weeks
Sample	Number of participants Grade/class standing Type of sample (K-12 or HE) Country
Higher-order, critical, or critical-analytic thinking measure	Reference Student-/Teacher-/Researcher-coded The score for the measure, based on predetermined criteria (1 = exclude, 2 = maybe, 3 = include)
Findings	The score for the type of findings (-1 = negative, 0 = no effect, 1 = positive) Specifics about the findings
Exclusion decision	Yes/no Official exclusion reason

HOT Higher-order thinking, *CT* Critical thinking, *CAT* Critical analytic thinking, *PBL* Problem-based learning, *PjBL* Project-based learning, *HE* Higher education

education setting and 6 in K-12. Twelve studies took place within the Health Sciences domain (e.g., nursing education, medical education), 8 within the Science, Technology, Engineering, Art, and Math (STEAM) domain, and 8 in other domains (e.g., financial management or psychology). Studies covered 12 different countries. Most studies were conducted in the USA ($n = 7$), Turkey ($n = 4$), Indonesia ($n = 4$), and China ($n = 3$). Two studies investigated the effects of PjBL; all other studies examined a PBL setting.

Research Question 1: Conceptualization of HOT, CT, and CAT

The first research question of this review was "How are HOT, CT, and CAT conceptualized in student-centered learning environments (i.e., PBL and PjBL)?" To answer this research question, it is important to know that none of the identified

Table 2 Descriptive Information of the Included Articles (*n* = 28)

Article	Treatment	Construct	Instrument	Duration	Level	Domain	Country
Burris and Garton (2007)	PBL	CT	WGCTA (Form S)	2 weeks	K-12	Other	USA
Carriger (2016)	PBL	CT	Researcher-developed assignment and rubric based on the Association of American Colleges and Universities	1 semester	HE	Other	USA
Choi et al. (2014)	PBL	CT	Critical Thinking Ability Scale for College Students	16 weeks	HE	Health	Korea
Cortázar et al. (2021)	PjBL	CT	Researcher-developed scale using a design-based research approach, starting with a CT definition from the American Philosophical Association	1 semester with 5×80 min classes	HE	STEAM	Chile
da Costa Carbogim et al. (2018)	PBL	CT	CCTDI and CCTST (Portuguese versions)	3 meetings (25 h)	HE	Health	Brazil
Dakabesi and Luoise (2019)	PBL	CT	Researcher-developed based on Ennis	6 meetings	K-12	STEAM	Indonesia
Dilek Eren and Akinoglu (2013)	PBL	CT	CCTDI (Turkish version)	14 weeks	HE	STEAM	Turkey
Ding (2016)	PBL	CT	CCTDI (researcher-adapted)	9 weeks	HE	Other	China
Fitriani et al. (2020)	PBL	CT	Critical Thinking Evaluation Package (Ennis)	5 months	K-12	STEAM	Indonesia
Gholami et al. (2016)	PBL	CT	CCTST (Form B)	8 weeks	HE	Health	Iran
Hassanpour Dehkordi and Heydarnejad (2008)	PBL	CT	Authors report to have used the CCTDI, but the results table includes the names of the subscales of CCTST	1 semester	HE	Health	Iran
Iwaoka et al. (2010)	PBL	CT	Cornell Critical Thinking Test	1 semester	HE	Other	USA
Lyons (2008)	PBL	CT	Assessment Technologies Institute (ATI) Critical Thinking Test	17 weeks (2 h per week)	HE	Health	USA
Muehlenkamp et al. (2015)	PBL	CT	Researcher-designed scales: Higher thinking, process, tools, and analysis	16 weeks	HE	Other	USA
Ozturk et al. (2008)	PBL	CT	CCTDI (Turkish version)	Curriculum approach	HE	Health	Turkey
Pardamean (2012)	PBL	CT	Health Sciences Reasoning Test (based on CCTST)	Curriculum approach	HE	Health	USA

Table 2 (continued)

Article	Treatment	Construct	Instrument	Duration	Level	Domain	Country
Rehmat and Hartley (2020)	PBL	CT	Test of Critical Thinking	16 weeks	K-12	STEAM	USA
Santiprasitkul et al. (2013)	PBL	CT	Critical Thinking Questionnaire	12 weeks (3 weeks per problem)	HE	Health	Thailand
Saputro et al. (2020)	PBL	CT	Critical Thinking Questionnaire (author-designed student survey based on Ennis)	6 weeks (810 min)	HE	STEAM	Indonesia
Sasson et al. (2018)	PjBL	CT &	Author-constructed test based on Dori et al. (2003) &	2-year program	K-12	Other	Israel
		HOT	Author-constructed test e.g., based on Bloom's taxonomy				
Şendağ and Odabaşı (2009)	PBL	CT	WGCTA (Form S; Turkish version)	8 weeks	HE	Other	Turkey
Siew and Mapeala (2016)	PBL	CT	Researchers constructed Test of Science Critical Thinking	18 h	K-12	STEAM	Malaysia
Sugeng and Suryani (2020)	PBL	HOT	Author-designed test based on Bloom's taxonomy	3 months (12 meetings)	HE	Other	Indonesia
Temel (2014)	PBL	CT	CCTDI (Turkish version)	5 sessions	HE	STEAM	Turkey
Tiwari et al. (2006)	PBL	CT	CCTDI	28 weeks	HE	Health	Hong Kong
D. Yu et al. (2013)	PBL	CT	CCTDI (Chinese version)	half a semester	HE	Health	China
W.-C. W. Yu et al. (2015)	PBL	CT	CCTDI	18 weeks	HE	Health	Taiwan
Yuan et al. (2008a)	PBL	CT	CCTST (Form A, Chinese-Taiwanese version)	18 weeks (36 h)	HE	Health	China

PBL Problem-based learning, *PjBL* Project-based learning, *CT* Critical thinking, *HOT* Higher-order thinking, *WGCTA* Watson–Glaser Critical Thinking Appraisal, *CCTDI* California Critical Thinking Disposition Inventory, *CCTST* California Critical Thinking Skills Test, *HE* Higher education, *STEAM* Science, Technology, Engineering, Art, and Math, *USA* United States of America

studies investigated the effect of P(j)BL on CAT. HOT was only investigated in two studies. One of the included studies investigated the effect of PBL on HOT, and the other study investigated the effect of PjBL on HOT and CT. All other studies investigated the effect of P(j)BL on CT. Hence, to answer this research question, we will mainly focus on the conceptualization of CT and, to a smaller extent, on HOT. In answering this research question, we will discuss how CT and HOT were defined and measured.

Conceptualization of CT

CT conceptualizations in P(j)BL consisted of CT dispositions and CT processes. These processes were referred to as "skills" or "abilities" in the included studies (e.g., W.-C. W. Yu et al., 2015). Our review will use the term "processes" instead of "skills." A CT disposition is "the constant internal motivation to engage problems and make decisions by using CT" (Facione et al., 2000, p. 65). Facione et al. (2000) use the word disposition to refer to individuals' characterological attributes. An example of a disposition is being open-minded, analytical, or truth-seeking. In the included studies, dispositions were described as the "will" or "inclination" to evaluate situations critically (e.g., Temel, 2014; W.-C. W. Yu et al., 2015) and a necessary pre-condition for CT processes (Temel, 2014).

Although the correlation between CT dispositions and processes is not extremely high (e.g., Facione et al., 2000), both seem to be necessary for "reasonable reflective thinking focused on deciding what to believe or do" (e.g., Ennis, 2011, p. 10). The included studies in this review defined the concept of CT seven times in terms of dispositions and processes, 13 times solely in terms of processes, and four times solely in terms of dispositions. Three studies did not define the concept or spoke in general terms as "a way to find meaning in the world in which we live" (Burris & Garton, 2007, p. 106). Of the included studies, 7 studies measured CT disposition, 18 studies measured CT processes, and one study measured CT dispositions and processes. One study stated to have measured CT dispositions, but in the results section, only statistics for CT processes were reported (Hassanpour Dekhordi & Heydarnejad, 2008). Regarding measurements, 16 studies were congruent in defining CT and measuring CT. This means that, for example, when authors defined CT in terms of dispositions, they measured dispositions as well. This also means that 11 studies were incongruent in this respect (e.g., mentioning processes but measuring dispositions or mentioning both but measuring one component).

In sum, although CT consists of dispositions and processes, in the conceptualization of CT in P(j)BL research, we saw a majority of studies focusing only on the processes and, to a lesser extent, on (the combination with) dispositions. Also, 11 studies were incongruent in focus (dispositions and/or processes) of their description of CT and the focus of their measurement instrument.

CT Dispositions In the studies included in this review, many different terms were used to describe CT disposition(s). The nine studies that reported measuring CT dispositions used the California Critical Thinking Disposition Inventory (CCTDI;

Facione, 1990a). The CCTDI assesses students' willingness or inclination toward engaging in critical thinking. The CCTDI contains seven dispositions (Facione, 1990a; Yeh, 2002). The scale *Truth-Seeking* refers to the mindset of being objective, honest, and seeking the truth even when findings do not support one's opinions/interests. *Open-Mindedness* refers to tolerance, an open mind toward conflicting views, and sensitivity toward the possibility of one's own bias. The subscale *Analyticity* concerns a disposition to anticipate possible consequences, results, and problematic situations. The fourth subscale, *Systematicity,* measures having an organized, orderly, and focused approach to problem-solving. The trust in one's reasoning process is measured in the subscale *CT Self-Confidence*. *Inquisitiveness* concerns intellectual curiosity, whereas *Cognitive Maturity* refers to the expectation of making timely, well-considered judgments. A qualitative analysis of the descriptions showed that most of the terms used corresponded with CCTDI subscales (Facione, 1990a; Yeh, 2002; see Appendix Table 5). Sometimes other terms were used that could be classified less easily according to the dispositions of the CCTDI but point toward a willingness to engage in critical thinking or the role of self-regulation and metacognition.

CT Processes There were not only many dispositions mentioned by the studies included in this review, but the number of terms to describe CT processes was even higher. When we look at the three most used instruments, many of these terms appear as specific components of those tests (see Appendix Table 6). For example, the most commonly used instruments were the California Critical Thinking Skills Test (CCTST; 4 studies) and the Watson–Glaser Critical Thinking Appraisal (WGCTA; 2 studies). One study used the Cornell Critical Thinking Test (CCTT), but the theoretical framework related to this test was used in three other studies to measure CT processes. The CCTST, WGCTA, and CCTT are well-known commercial standardized measures of critical thinking.

The CCTST is a companion test of the CCTDI and measures five critical thinking processes: Analysis, Evaluation, Inference, Deductive Reasoning, and Inductive Reasoning (Facione, 1991). *Analysis* refers to accurately identifying problems and processes such as categorization, decoding significance, and clarifying meaning. *Evaluation* concerns the ability to assess statements' credibility and arguments' strength. The *Inference* subscale measures the ability to draw logical and justifiable conclusions from evidence and reasons. *Deductive Reasoning* relies on strict rules and logic, such as determining the consequences of a given set of rules, conditions, principles, or procedures (e.g., syllogisms, mathematical induction). Finally, *Inductive Reasoning* refers to reasoned judgment in uncertain, risky, or ambiguous contexts.

Another well-validated test used in the studies included in this review is the WGCTA (Watson & Glaser, 1980). The WGCTA provides problems and situations requiring CT abilities. It measures CT as a composite of attitudes of inquiry (i.e., recognizing the existence of problems and acceptance of a need for evidence), knowledge (i.e., about valid inferences, abstractions, and generalizations), and skills in applying these attitudes e and knowledge (Watson & Glaser, 1980, 1994, 2009).

The scale consists of five subscales: Inferences, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. The *Inferences* subscale measures to what extent participants can determine the truthfulness of inferences drawn from given data. *Recognition of Assumptions* concerns recognizing implicit presuppositions or assumptions in statements or assertions. The *Deduction* subscale measures the ability to determine if conclusions necessarily follow from the given information. *Interpretation* concerns weighing evidence and deciding if the generalizations based on the given data are justifiable. Finally, the subscale *Evaluation of Arguments* measures the ability to distinguish strong and relevant arguments from weak and irrelevant arguments. The long version of the scale consists of 80 items (parallel Forms A and B; Watson & Glaser, 1980), and the short version (From S) contains 40 items (Watson & Glaser, 1994). Newer test versions are available (Watson & Glaser, 2009), but the included studies relied on the older, abbreviated version (Burris & Garton, 2007; Şendağ & Odabaşı, 2009).

The CCTT level Z, designed by Ennis, measures deduction, semantics, credibility, induction, definition and assumption identification, and assumption identification (Bataineh & Zghoul, 2006; Ennis, 1993). The CCTT is a commercial measure of critical thinking and has two versions. The CCTT level X is for students in Grades 4–14, whereas CCTT level Z is for advanced and gifted high school students, college students, graduate students, and other adults. Possibly due to copyright restrictions, many articles gave very brief descriptions of the test subscales. We also found some inconsistencies in the descriptions of the subscales in the literature. We could not retrieve the original manuals of the CCTT from 1985 or the revised version from 2005. Leach and others (2020) gave a detailed description of the CCTT level X subscales and provided a possible explanation for the inconsistencies found in the literature.

The test measures five latent dimensions: Induction, Deduction, Observation, Credibility, and Assumption (Leach et al., 2020). However, these five dimensions are reduced into four parts in the test manual. Two dimensions are taken together (i.e., observation and credibility), and some items of one dimension are counted as an element of another part (Leach et al., 2020). Bataineh and Zghoul (2006) described the subscales of CCTT level Z in more detail. According to them, the CCTT level Z measures six dimensions: Deduction, Semantics, Credibility Induction, Definition and Assumption Identification, and Assumption Identification. The subscale *Deduction* measures to what extent a person can detect valid reasoning. The subscale, *Semantics,* measures the ability to assess verbal and linguistic aspects of arguments. *Credibility* concerns the extent to which a participant can estimate the truthfulness of a statement. The subscale *Induction* refers to the ability to judge conclusions and the best possible predictions. *Definition and Assumption Identification* measures the extent to which a person can identify the best definition of a given situation. Finally, *assumption identification* asks participants to choose the most probable unstated assumption in the text.

In summary, as can be seen from this descriptive analysis (see Appendix Tables 5 and 6), many terms are employed to characterize CT dispositions and processes. The conceptualization of CT differs per measurement, as evidenced by the three most

Table 4. Main Results of the Included Studies ($n = 2$) on Higher-Order Thinking (HOT)

Authors (year)	Treatment	Design	Sample size		Main findings
			P(j)BL)	Control	
Sasson et al. (2018)	PjBL	IG pre- 2 post	38	27	HOT increased for the PjBL group but not for the control group from Measurement 1 (beginning 9th grade) to 3 (end of 10th grade; dependent *t*-tests)
Sugeng and Suryani (2020)	PBL	IG	48	?	PBL scored significantly higher on HOT, whereas the control group scored higher on lower-order-thinking

PBL Problem-based learning, *PjBL* Project-based learning, *IG* Independent groups

commonly used instruments (i.e., CCTDT, CCTST, WGCTA) and the instruments based on Ennis's conceptualization of CT (e.g., CCTT). We also observed a tendency to create new instruments, often based on or inspired by other measurement instruments that introduce new terms to describe CT.

HOT Processes

In some studies measuring CT, the authors mentioned that CT was a component of HOT (Cortázar et al., 2021; Dakabesi & Louise, 2019; Sasson et al., 2018). Only two studies measured HOT (Sasson et al., 2018; Sugeng & Suryani, 2020). In both instances, the authors solely defined HOT in reference to the more complex thinking levels in Bloom et al. and's (1956; Krathwohl, 2002) taxonomy for the cognitive domain: application, analysis, synthesis, and evaluation. In the Sasson et al. (2018) study, comprehension was also included as one of the higher cognitive processes. In contrast, in the Sugeng and Suryani (2020) study, comprehension was treated as a form of lower-order thinking. Both studies use Bloom's Taxonomy as a framework for coding student work. Therefore, we mainly saw the same terms to conceptualize HOT and also congruency between characterizing and measuring HOT processes, based on Bloom's taxonomy.

Research Question 2: Can PBL and PjBL Foster HOT, CT, and CAT?

To answer Research Question 2, we summarized the main findings regarding the effectiveness of PBL and PjBL on CT and HOT (see Tables 3 and 4). When possible, we calculated the standardized mean difference (Cohen's *d*) in Comprehensive Meta-Analysis statistical software (version 3; Borenstein et al., 2009). First, we will discuss the effects on CT and then the two studies that examined HOT.

Table 3. Main Results of the Included Studies ($n = 27$) on Critical Thinking

Authors (year)	Treatment	Design	Sample size P(j)BL	Sample size Control	Main findings	Effect direction	Cohen's d [95% CI]
Burris and Garton (2007)	PBL	IG pre-post	77	63	CT total score: Control group (supervised study) scored significantly higher than the PBL group on the CT posttest (ANCOVA, pretest as covariate). Subscale scores: not reported	-	-0.02 [-0.36, 0.31]
Carriger (2016)	PBL	IG	32	18	CT score: PBL scored significantly higher than the control group (ANOVA)	+	–
Choi et al. (2014)	PBL	IG pre-post	46	44	CT total score: No significant difference between PBL and control groups on the CT posttest (ANCOVA, pretest as covariate). Subscale scores: not reported	ns	0.26 [-0.16, 0.67]
Cortázar et al. (2021)	PjBL	Pre-post	266	–	CT score: PjBL students increased from pretest to posttest	+	0.05 [-0.04, 0.15]
da Costa Carbogim et al. (2018)	PBL	Pre-post	50	–	CT total scores: No significant differences for the CCTDI and CCTST total scores from pretest to posttest (dependent t-test) Subscale scores: No significant differences on the subscales of the CCTDI and CCTST	ns	-0.05 [-0.33, 0.23]
Dakabesi & Louise (2019)	PBL	IG	60	64	CT score: PBL group scored significantly higher on the posttest than the control group (independent t-test)	+	0.42 [0.06, 0.78]

Table 3. (continued)

Authors (year)	Treatment	Design	Sample size		Main findings	Effect direction	Cohen's d [95% CI]
			P(j)BL	Control			
Dilek Eren et al. (2013)	PBL	IG pre-post	24	22	CT total score: No significant difference on the pretest. Higher score for PBL on the posttest (independent t-test). Subscale scores: No significant differences on the pretest. Higher score for PBL on the posttest scores of Inquisitiveness and Systematicity	+	1.14 [0.51, 1.76]
Ding (2016)	PBL	Pre-post	316	–	CT total score: not reported. Subscale scores: Significant increases on all CCTDI scales	+	0.45 [0.36, 0.54]
Fitriani et al (2020)	PBL	IG pre-post	36	32	CT total score: PBL had a higher gain score than the control group (ANOVA). Subscale scores: Not discussed if there were significant differences	+	–
Gholami et al. (2016)	PBL	IG pre-post	40	40	CT total score: PBL group showed a higher increase from pretest to posttest (t-test on gain scores). Subscale score: PBL showed a higher increase from pretest to posttest for Analysis	+	0.43 [-0.01, 0.88]
Hassanpour Dehkordi and Heydarnejad (2008)	PBL	IG pre-post	20	20	CT total score: PBL scored higher than the control group (statistics not provided). Subscale scores: Only statistics for the scales Analysis, Inference, and Evaluation are reported, showing higher posttest scores for PBL	+	1.45 [0.75, 2.15]

Table 3. (continued)

Authors (year)	Treatment	Design	Sample size		Main findings	Effect direction	Cohen's d [95% CI]
			P(j)BL	Control			
Iwaoka et al. (2010)	PBL	Pre-post	154	–	CT score: No significant gain from pretest to posttest across 8 course implementations (2 out of 8 implementations showed a significant increase; dependent t-tests). Subscale score: No significant gains	ns	0.10 [-0.06, 0.26]
Lyons (2008)	PBL	IG pre-post	27	27	CT score: No significant differences between PBL and control groups on the CT posttest (ANCOVA, pretest as covariate)	ns	0.01 [-0.52, 0.55]
Muehlenkamp et al. (2015)[a]	PBL	IG and pre-post	38	36	CT total score: not reported. Subscale scores: The PBL group showed a significant gain from pretest to posttest on all scales (Higher Thinking, Process, Tools, and Analysis). On the posttest, PBL students scored higher than the control group on all scales	+	1.62 [1.09, 2.15]
Ozturk et al. (2008)	PBL	IG	52	95	CT total score: PBL group scored significantly higher than the control group (independent t-test). Subscale scores: Significant differences in favor of PBL for Open-Mindedness, Truth-Seeking, and Systematicity	+	0.43 [0.09, 0.77]

Table 3. (continued)

Authors (year)	Treatment	Design	Sample size P(j)BL	Sample size Control	Main findings	Effect direction	Cohen's d [95% CI]
Pardamean (2012)	PBL	IG (3 PBL years)	98	–	CT total score: No difference between first-, second-, and third-year students (ANOVA). Subscale scores: Significant difference between the 3-year groups for Inductive Reasoning (highest score for second-year students, lowest score for third-year students)	ns	–
Santiprasitkul et al. (2013)	PBL	Pre-post	94	–	CT total score: Significant increase from pretest to posttest. Subscale scores: Significant pretest–posttest gains on 2 (Inference and Recognition of Assumption) of the 5 CT subscales (dependent t-tests)	+	0.50 [0.33, 0.67]
Saputro et al. (2020)	PBL	IG pre-post	22	22	CT score: No significant difference on the pretest. Higher score for PBL on the posttest (independent t-test)	+	3.43 [2.51, 4.36]
Sasson et al. (2018)[b]	PjBL	IG pre- 2 post	38	27	CT score: No significant differences between the PjBL and the control group on the pretest (beginning 9th grade) and posttest (end 9th grade). PBL scored significantly higher on the second posttest (end of 10th grade; independent t-tests)	+	0.86 [0.34, 1.37]

Table 3. (continued)

Authors (year)	Treatment	Design	Sample size		Main findings	Effect direction	Cohen's *d* [95% CI]
			P(j)BL	Control			
Şendağ and Odabaşı (2009)	PBL	IG pre-post	20	20	CT total score: Significant interaction (time x treatment), indicating a higher increase for PBL (mixed ANOVA). Subscale scores: not reported	+	0.66 [0.02, 1.29]
Siew and Mapeala (2016)	PBL	IG pre-post	90	90	Subscale scores: PBL had a higher increase than the control group on 2 of the 3 measures	+	0.58 [0.28, 0.88]
Rehmat and Hartley (2020)	PBL	IG pre-post	52	46	CT score: Significant interaction between time and teaching method indicating a higher increase for PBL (mixed ANOVA)	+	2.81 [2.25, 3.37]
Temel (2014)	PBL	IG pre-post	22	27	CT total score: No significant interaction (time x treatment) or a significant increase from pretest to posttest (ANCOVA, pretest as covariate). Subscale scores: not reported	ns	0.14 [-0.43, 0.70]
Tiwari et al. (2006)[c]	PBL	IG pre-post+2 follow-ups	38	36	CT total score: Higher increase for PBL during the course; PBL group continued to have higher (gain) score one year later (*t*-test on gain scores). Subscale scores: Higher increases for PBL during the course for CT Self-Confidence (gain score 1), Analyticity (gain scores 1 and 2), and Truth-Seeking (gain scores 1–3)	+	0.69 [0.22, 1.16]

Table 3. (continued)

Authors (year)	Treatment	Design	Sample size		Main findings	Effect direction	Cohen's d [95% CI]
			P(j)BL	Control			
D. Yu et al. (2013)[d]	PBL	IG	37	39	CT total score: PBL group scored significantly higher than the control group (independent t-test). Subscale scores: Significant differences in favor of PBL for Open-Mindedness and Inquisitiveness	+	0.51 [0.05, 0.96]
W.-C. W. Yu et al. (2015)	PBL	IG pre-post	40	41	CT total score: No significant differences between groups on CT on the pretest and posttest (MANOVA). Subscale scores: No significant differences between groups on CT subscales on the pretest and posttest	ns	–
Yuan et al. (2008a)	PBL	IG pre-post	23	23	CT total score: Higher gain score PBL than the control group (t-test on gain scores). Subscale scores: Significant differences in gain scores favoring PBL for Analysis and Induction	+	0.83 [0.23, 1.43]

PBL Problem-based learning, *PjBL* Project-based learning, *IG* Independent groups, *ANCOVA* Analysis of Covariance, *(M)ANOVA* (Multivariate) Analysis of Variance, *CCTDI* California Critical Thinking Disposition Inventory, *CCTST* California Critical Thinking Skills Test

[a]For the effect size calculation, we used the posttest scores of both groups

[b]The effect size was based on the first and second posttest

[c]The effect size was based on the pretest and first posttest scores

[d]The semester was split into two learning processes; during the first half, one of the groups received PBL, whereas the other group received lecture-based learning; in the second learning process, this was reversed (cross-over experimental design). The main findings and effect size calculation were based on the posttest scores of the first-PBL (treatment) and first-lecture-based (control) groups

Effects of P(j)BL on Critical Thinking

Description of Studies Table 3 reports the main findings of the 27 studies that investigated the effects of P(j)BL on CT. Most studies investigated the effects of PBL, and only two examined the effects of PjBL (Cortázar et al., 2021; Sasson et al., 2018). Most studies compared P(j)BL with a control group with pre-post measures of CT ($n=16$). Five studies compared PBL with a control group and only reported posttest scores (see Table 3). We used the pre-post data of a single P(j)BL group (without a control group) for five other studies. One study had another design (comparing year groups; Pardamean, 2012).

The most commonly reported control group was (traditional) lecture-based learning (e.g., Carriger, 2016; Choi et al., 2014; Gholami et al., 2016; Lyons, 2008; Rehmat & Hartley, 2020; Tiwari et al., 2006; W.-C. W. Yu et al., 2015; Yuan et al., 2008a), traditional or conventional learning (e.g., Dilek Eren & Akinoglu, 2013; Fitriani et al., 2020; Saputro et al., 2020; Sasson et al., 2018; Temel, 2014), or instructor-led instruction (Şendağ & Odabaşı, 2009). In Burris and Garton (2007), the control group was a supervised study group, of which the authors indicated this corresponded with Missouri's recommended curriculum for (secondary) agriculture classes. The control group in Siew and Mapeala (2016) focused on conventional problem-solving.

Some studies included additional experimental groups (Carriger, 2016; Cortázar et al., 2021; da Costa Carbogim et al., 2018; Fitriani et al., 2020; Siew & Mapeala, 2016; W.-C. W. Yu et al., 2015). In these experimental groups, PBL was combined with another intervention (Cortázar et al., 2021; da Costa Carbogim et al., 2018; Fitriani et al., 2020; Siew & Mapeala, 2016; W.-C. W. Yu et al., 2015) or combined with lectures (Carriger, 2016). For example, these interventions might include a critical thinking intervention (da Costa Carbogim et al., 2018) or a socially shared regulation intervention (Cortázar et al., 2021). To describe the main findings and calculate effect sizes, we only report the data for the "regular" PBL and the control groups (if reported) of these studies. Overall, results showed positive effects of P(j)BL on critical thinking. When we only look at the statistical tests the authors performed, 19 studies reported positive effects on CT, indicating that students' CT disposition or skills scores increased from pretest to posttest or obtained higher scores than the control group. Seven studies reported non-significant findings, and only one study reported a negative effect.

Meta-Analysis In the meta-analysis section of this review, we only included studies with a pre-post and/or independent groups design. Independent groups designs and pre-post designs both give insight into the question of whether PBL and PjBL affect students' CT. Independent group designs check if the instructional method is more effective than "traditional" education, whereas pre-post designs check for differences before and after the implementation.

We were able to calculate effect sizes for 23 studies. For the studies with independent groups pre-post designs, we used the pretest and posttest means and standard deviations (*SD*s) and sample size per group to compute effect sizes. Posttest *SD* was used to standardize the effect size. Because most studies did not report the correlation between the pretest and posttest scores, we assumed a conservative correlation of 0.70 if the correlation was not reported. Studies suggest strong test–retest correlations for standardized
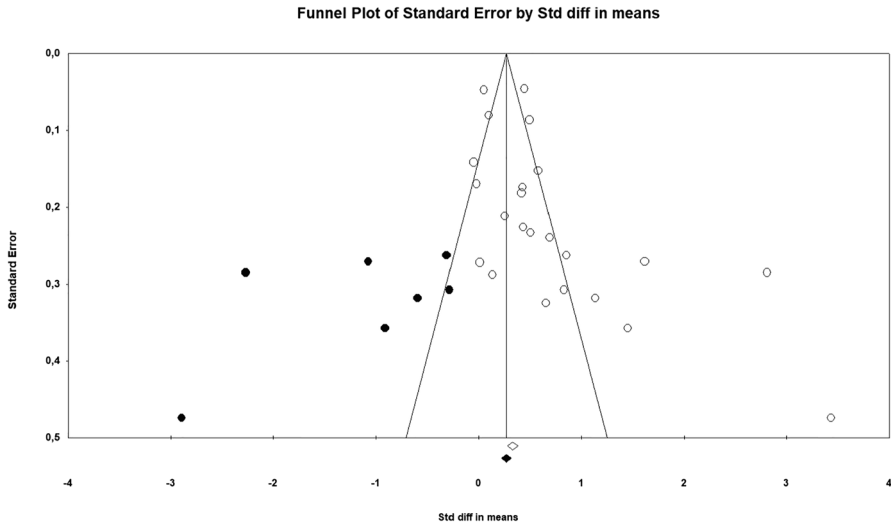
Fig. 2 Funnel Plot. *Note.* Funnel plot with observed and imputed studies. The white dots represent the observed study samples included in the meta-analysis; the black dots represent the seven studies trimmed at the left side using Duval and Tweedie's trim-and-fill technique

critical thinking measures (Gholami et al., 2016; Macpherson & Owen, 2010). For example, Macpherson and Owen (2010) reported a strong positive correlation ($r = 0.71$) between two test moments among medical students for WGCTA. We used both groups' means, *SD*s, and sample sizes to calculate the effect sizes for the studies with independent groups posttest-only designs. We used the mean difference, *t*, and sample size for the single group pre-post studies to calculate the effect size (da Costa Carbogim et al., 2018; Iwaoka et al., 2010) or the pretest and posttest means and SDs, sample size, and pre-post correlation. Again, we assumed a correlation of 0.70 if it was not reported.

When we included all 23 studies in one analysis (random-effects model), this resulted in a medium effect size of 0.644 ($SE = 0.10$, 95% CI [0.45, 0.83]). These results suggest that P(j)BL could positively affect students' CT dispositions and processes. However, the effect was heterogenous, $Q(22) = 237.46$, $p < 0.001$, $I^2 = 90.74$, $T^2 = 0.17$ ($SE = 0.11$), which implies that the variability in effect sizes has sources other than sampling error. Because only one of the studies investigated PjBL, we repeated the analysis for only the PBL studies, resulting in an effect size of 0.635 ($SE = 0.10$, 95% CI [0.44, 0.83]), indicating the results remained similar. In our analysis, we included three types of research designs. Overall, the studies that compared P(j)BL with a control group reported larger effect sizes (independent groups: $n = 4$, $d = 0707$, $SE = 0.22$; independent groups pre-posttest: $n = 14$, $d = 0.831$, $SE = 0.13$). than studies with a single-group pre-post design ($n = 5$, $d = 0.213$, $SE = 0.18$).

As Table 3 reveals, there were some studies with extreme effect sizes (e.g., Saputro et al., 2020). We, therefore, conducted leave-one-out analyses. The leave-one-out analyses revealed that effect sizes were between 0.533–0.686 with an *SE* of approximately 0.10. Confidence intervals were between 95% CI [0.37, 0.70] to 95% CI

[0.48, 0.89]. Overall, the conclusion about the positive effect of P(j)BL on critical thinking would not change if we left out the study by Saputro et al. (2020).

We further examined publication bias by inspecting the funnel plot and Egger's reception intercept (Egger et al., 1997). We applied Duval and Tweedie's (2000) trim-and-fill technique and conducted a classic fail-safe N analysis. Figure 2 presents the funnel plot of all included studies and plots the individual study effect size against the standard error of the effect size estimates. The funnel plot indicated publication bias. Egger's linear regression test for asymmetry further supported this observation, $t(21) = 2.78$, $p = 0.006$. Duval and Tweedie's trim-and-fill technique (7 studies trimmed at the left side) resulted in an adjusted effect size from a medium effect of 0.644 to a small effect of 0.298 (95% CI [0.09, 0.51]). The fail-safe N suggested that 1,376 missing studies are needed for the result of this meta-analysis to be nonsignificant ($p > 0.05$). Overall, the results suggested that P(j)BL can have a small-to-medium positive effect on students' CT processes and dispositions.

As mentioned, the effect was heterogeneous. Due to the limited number of studies, we could not investigate moderating factors that can explain the heterogeneity statistically. The variation in effect sizes can likely be partly caused by variation in how PBL was implemented, as this often differs per institute even when the defining characteristics have been met (Maudsley, 1999; Norman & Schmidt, 2000). However, to deal with this issue, we only included studies that met the defining criteria of PBL and PjBL. We further excluded studies that contained additional activities (e.g., concept mapping) that could affect the results.

Also, differences in the exact operationalization of CT could affect the results. To explore this, we calculated the effect size separately for studies reporting outcomes on CT processes ($n = 17$) and dispositions ($n = 7$). Analyses suggested a higher effect size for the studies reporting results for CT processes ($d = 0.720$, $SE = 0.14$, 95% CI [0.46, 0.99]) than studies reporting on the effects of CT disposition ($d = 0.411$, $SE = 0.13$, 95% CI [0.16, 0.66]). However, these results must be interpreted with caution due to the limited number of studies and the extreme effect sizes in the CT processes group.

Other variables that could potentially explain heterogeneity are sample level (e.g., K-12 or higher education) and the duration of the intervention or exposure to PBL. For example, a meta-analysis of the effects of student-centered learning on students' motivation showed that the effect of student-centered learning on motivation was lower for K-12 samples and curriculum implementation compared to studies conducted in a higher education setting and course implementations (Authors, 2022). Possibly, similar factors could affect the effect of P(j)BL on CT, but more research is needed to investigate this in more detail.

Additional Findings In the meta-analysis section of this review, we only included studies with a pre-post and/or independent groups design. Two studies deviated from this design. Tiwari et al. (2006) did not only compare the effects of PBL vs. lecture-based learning immediately after the PBL course but also included two follow-ups one and two years later. As seen in Table 3, the PBL group showed significant gain scores in CT immediately after the course, and the gain score remained positive at the first follow-up. However, the gain score became non-significant at the second follow-up (two years later). When we look at some subscale scores, results revealed significant gains in favor of PBL for truth-seeking, analyticity, and CT self-confidence. The gain

score for analyticity remained positive at the first follow-up, whereas the truth-seeking gain score remained significant at the first and second follow-ups (two years later). The results of this study suggest that PBL can have long-term effects on CT dispositions.

Pardamean (2010) did not have a control group but examined CT processes in first through third-year students. Their study revealed no differences between the three-year groups on overall CT. There was one statistically significant difference on the subscale Inductive Reasoning, on which the second-year students obtained the highest score and the third-year students the lowest score. This study does not support that CT increases across year groups in a PBL curriculum. However, we have no information on the baseline CT of each group.

Not all studies reported the results of the test subscales, even when the original scale consisted of subscales. Seven of the nine studies using the CCTDI reported subscale results. Of these studies, three studies reported positive results for Open-Mindedness, Inquisitiveness, Truth-Seeking, or Systematicity, and two for the Analyticity subscale or the CT Self-Confidence subscale. Of the five studies that examined Cognitive Maturity, only one reported a positive effect of P(j)BL (see Table 3). Five studies reported the subscale scores on the CCTST. Of these studies, two found positive effects of P(j)BL on Analysis or Evaluation, and one study on Inference. No students reported positive effects on the Deduction subscale. For Induction, one study found a positive effect, and another reported a negative effect. Overall, mixed results were found on the subscale level (see Table 3).

Effects of P(j)BL on Higher-Order Thinking

Only two studies investigated the effects of P(j)BL on HOT (see Table 4). One in a PBL setting (Sugeng & Suryani, 2020) and the other in a PjBL setting (Sasson et al., 2018). We could not calculate effect sizes based on the data provided in the papers. Sugeng and Suryani (2020) compared a PBL group with a lecture-based group on HOT and lower-order thinking. The PBL group scored significantly higher on HOT, whereas the lecture-based group scored higher on lower-order thinking. Sasson et al. (2018) reported a positive effect for a 2-year PjBL program. HOT increased for the PjBL group but not for the control group from Measurement 1 (beginning of 9th grade) to Measurement 3 (end of 10th grade).

Conclusions and Implications

This systematic review focused on two questions: "How are HOT, CT, and CAT conceptualized in student-centered learning environments?" and "Can PBL and PjBL foster HOT, CT, and CAT?" We presented and discussed findings related to those questions in the preceding section. Here we offer a more global examination of the trends that emerged from our analysis of two popular forms of student-centered approaches to instruction, PBL and PjBL, and share lingering issues that should be explored in future research. However, we first address certain limitations of this systematic review that warrant consideration.

Limitations

As stated, several limitations emerged in this systematic review that have a bearing on the conclusions we proffer. For one, when we set out to conduct this research, our intention was to understand how higher-order, critical, and critical-analytic thinking were conceptualized in PBL and PjBL. However, we found it impossible to analyze the role of CAT because CAT was not part of any investigation in the context of P(j) BL. Further, the attention given to HOT was quite limited, with only two studies investigating it. It should also be mentioned that we found an unequal distribution in the studies included in this review regarding the thinking measures and the learning environments. For example, CAT appeared not embedded in the P(j)BL literature and HOT to a significantly smaller degree than CT.

Similarly, most studies in this review reported findings of a PBL environment, with a PjBL environment only investigated in two studies. Consequently, the conclusions we draw from our analysis rest primarily on empirical research on studies applying PBL and not PjBL approaches.

As mentioned, the CT measures used in this work often consisted of several subscales. For example, the WGCTA consists of the subscales inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. The presence of multiple indicators was a limitation because the effects can differ for these subcomponents, making global interpretation of effectiveness more difficult. However, we did not use those components as search terms in the literature search, and most studies did not report these subscales or define them. Future studies could use finer-grained search terms, including the thinking measures' subscales or processes.

While the present study demonstrated positive effects of P(j)BL on HOT and CT, it remains unknown what exactly led to these positive effects, given that multiple links between P(j)BL and HOT and CT could be identified. Also, associations of HOT, CT, and CAT with performance were not investigated in the present study. More controlled experimental studies could shed light on these issues and help overcome the design issues associated with effect studies.

Finally, future research could relate HOT, CT, and CAT in P(j)BL environments to other learning processes, such as self-regulated learning (SRL) and self-directed learning (SDL). Components such as metacognition also play a prominent role in SRL and SDL processes. Future research could shed light on the relationships between thinking and regulating processes in the context of P(j)BL.

Research Question 1: CAT Is Not Embedded and HOT Not Frequently Studied in the P(j)BL Literature

Concerning conceptualizations (RQ1), we must first acknowledge the skewed distribution of studies over the three types of thinking (i.e., HOT, CT, and CAT). To start with, CAT was not part of any investigation in the context of P(j)BL, HOT only in two studies, with the vast majority focusing on CT. When looking at the definition of CAT and its distinguishing feature compared to CT, it has been put forward as the focus on determining how appropriate and credible evidence is (Byrnes & Dunbar, 2014). Remarkably,

this component is undoubtedly present in P(j)BL. After all, in PBL, when students work on the problem (that they encounter based on prior knowledge) or, more specifically, the learning questions/issues for further self-directed study formulated during PBL group discussion, they will look for and study different literature resources (e.g., Loyens et al., 2012). During this knowledge acquisition process in finding answers to the learning questions, they need to check whether different literature resources are in accordance with each other or whether dissimilarities can be detected. In case of dissimilarities, it is up to the student to decide and, later on, during the reporting phase, to discuss how to deal with these dissimilarities with the group. How come different sources provide different answers to the learning questions/issues, and what does that say about the credibility of the sources themselves? Also, in PjBL, students undergo the same process when dealing with conflicting information while working on their projects. It is important to note that these conflicting pieces of information are resolved through group discussion in P(j)BL. However, initially, they might cause some uncertainty regarding the learning process. Indeed, several scholars acknowledge learning uncertainty as a potential consequence of the open set-up of student-centered learning environments (Dahlgren & Dahlgren, 2002; Kivela & Kivela, 2005; Llyod-Jones & Hak, 2004). Nevertheless, the four steps described by Alexander (2014) are present in P(j)BL, which probably implies that the concept of CAT is not yet well known and embedded in the P(j)BL literature and that in the context of the P(j)BL literature, CAT would be a more accurate term compared to CT.

Similarly, only two studies examined HOT, which could be explained by the fact that HOT is an umbrella term consisting of CT (Schraw et al., 2011). That means that when researchers investigate HOT, they are also investigating CT. In light of conceptual clarity, however, it would be recommended to examine concepts at the most detailed level.

More Focus on CT Processes Than CT Dispositions

Another finding was that in the P(j)BL literature, more focus lies on processes, referred to as skills or abilities, compared to dispositions of CT or combinations of CT processes and dispositions. This is not surprising as P(j)BL has been more focused on and related to several interpersonal and self-directed learning skills (Loyens et al., 2008, 2012; Schmidt, 2000). What is more problematic is the incongruence between the definitions (CT dispositions and/or processes) and measurement instruments. From an educational point of view, processes seem to be the most natural to be fostered in education, although research has also demonstrated that learning environments can foster CT dispositions (Mathews & Lowe, 2011). Applied to P(j)BL environments, dispositions such as "inquisitiveness," "open-mindedness," "analyticity," and "self-regulatory judgment" are certainly helpful. However, more empirical research is necessary to see whether and how P(j)BL environments can foster CT dispositions. The exact meaning of dispositions, which are frequently measured in this literature, is unclear. The measures, subscales, or items labeled as dispositions range from rather stable personality traits such as open-mindedness to more malleable individual differences factors such as prior knowledge. Some components included under dispositions also carry a strong cognitive character, such as analyticity.

Lack of Conceptual Clarity Troubles Measurements and Findings

Regarding the conceptualizations of HOT, CT, and CAT, we definitely ended up in muddy waters. The largest percentage of excluded articles was due to flawed conceptualizations (also for P(j)BL, which we will explain below). These flawed conceptualizations produce a domino effect because conceptualizations (i.e., theoretical definitions) are determinative for measurements (Byrnes & Dunar, 2014). In addition, conceptualizations were, so we observed, often at best operationalizations in which authors named specific (sub)processes without mentioning any theoretical grounding. The measurement tool used was often determinative for the inclusion of specific processes. However, we often observed a mismatch in that processes were mentioned that were not part of the measurement instrument, or we observed incongruence between definitions (e.g., processes) and measurement instruments (e.g., measuring dispositions).

Of course, attempts have been made to reach a consensus regarding conceptualizations. For example, the Delphi Report (Faccione, 1990b) and the special issue on CAT (e.g., Alexander, 2014; Byrnes & Dunar, 2014) tried to create conceptual clarity on respectively CT and CAT. Consensus should, however, not be a goal in itself. HOT, CT, and CAT are such broad concepts that consensus is far from easy. It is, however, more important to reflect on what the learning objectives of the learning environment (P(j)BL or other) are and determine whether fostering HOT, CT, and CAT is one of them, and then create an educational practice that is in line with these objectives (i.e., constructive alignment; Biggs, 1996). Depending on the learning objectives of a learning environment (i.e., the construction of flexible knowledge bases, the development of inquiry skills, or a tool for "learning how to learn"; Schmidt et al., 2009), one could emphasize specific (sub)CT processes and use different measurement instruments.

A consequence of these flawed conceptualizations is flawed measurements. Measurements were often problematic in terms of their psychometric properties. Another issue in this respect is that many CT measurements are commercial and not readily available.

Next to issues with the conceptualizations of HOT, CT, and CAT, also serious problems were seen in the conceptualization of P(j)BL. Descriptions were absent, unclear, too broad and general (e.g., "active learning"), or indicative of a learning environment other than P(j)BL. This is not a new finding. In fact, PBL, for example, was identified as troublesome in terms of its definitions a long time ago (e.g., Lloyd-Jones et al., 1998). However, it is at least troublesome that this is still the case and, therefore, we excluded many studies. On the other hand, we also notice that more recent studies pay more attention to this issue (e.g., Lombardi et al., 2022).

Design Issues in Effect Studies of P(j)BL on HOT, CT, and CAT

While investigating the second research question of this study on the effects of P(j)BL on HOT, CT, and CAT, we made several observations. First, there is still a lack of controlled studies in this domain. The great majority of studies did not use a

control group (note: those included in the meta-analysis did), making it impossible to determine the effects of interventions. Like the unclear conceptualizations of P(j)BL, this is not new and has been indicated as an issue before (Loyens et al., 2012).

Another observation was that in many studies, it was seen as an assumption/given that P(j)BL fosters CT, usually without any explanation. As explained in the introduction, fostering HOT, CT, or CAT skills is not mentioned as one of the goals of P(j)BL, despite the links that can be made. A priori stating that P(j)BL fosters CT is hence premature.

Finally, we noticed that in several studies, PBL was combined with other interventions (e.g., concept mapping). In those cases, the PBL group served as a control group and the PBL plus extra group as the experimental condition, making it impossible to establish the effects of PBL on HOT, CT, and CAT. Most studies used a pre-posttest design.

Research Question 2: Positive effects of P(j)BL on HOT and CT

Overall, results showed positive effects of P(j)BL on CT and HOT (note that no studies on CAT were found to be included in this review), with scores increasing from pre- to posttest or P(j)BL obtaining higher scores than the control group. These findings imply that P(j)BL does carry elements that can foster CT and HOT. As mentioned above, the positive effects could be interpreted by both cognitive and instructional science literature. Literature on how to effectively teach HOT, CT, and CAT has identified several techniques that help foster these skills. Not surprisingly, these techniques, such as dealing with real cases/problems in class, encouraging (Socratic) class discussions/debate, fostering inquiry-oriented experiments, problem-solving, problem finding, brainstorming, decision making, and analysis (Abrami et al., 2015; Miri et al., 2007; Torff, 2011) are all linked to P(j)BL. Similarly, literature on the learning processes involved in P(j)BL also mentions processes linked to thinking skills. For example, students working on their projects in PjBL use problem-solving, design, decision-making, argumentation, using and weighing different pieces of knowledge, explanation, investigation, and modeling (Krajcik et al., 2008). Similarly, students working on problems in PBL have group discussions about authentic problems, engage in evidence-seeking behavior, analyze the evidence, resolve unclarities, and decide on the outcome.

While the majority of findings revealed positive effects regarding the effectiveness of P(j)BL in fostering CT and HOT, several studies also reported negative or no effects. Given the wide variety of P(j)BL formats, it might be due to implementation issues, but it can also be ascribed to design issues, as mentioned above. Further research needs to shed light on the null and negative findings.

As a final note, it should be mentioned that effects were found for CT, but the studies usually do not make claims about the finer-grained subprocesses. For example, four components of HOT have been identified (Schraw et al., 2011), while outcome measures are usually calculated at the "general" and not the subcomponent level. Nevertheless, the HOT component of metacognition is quite different from the HOT component of reasoning or problem-solving, which demonstrates the importance of constructive alignment (Biggs, 1996). Constructive alignment implies clearly defining the learning objectives of the learning environment (P(j)BL

or other). When fostering HOT, CT, and CAT is one of the learning objectives, an educational practice should be developed that aligns with these objectives. To make claims about whether learning environments are effective in fostering HOT, CT, or CAT processes, one must first discover whether these processes are or can be part of the learning objectives of these learning environments.

Implications

Several implications for theory and practice follow out of this review study. The first implication is that there is much room for improvement in terms of conceptual clarity. CAT is not investigated and HOT only sporadically in the context of P(j)BL. Nevertheless, when looking at the respective definitions of these thinking processes, the "analytical" part of CAT is certainly present in P(j)BL environments when weighing the evidence during the analysis of specific problems or projects (Alexander, 2014). In addition, since CT is considered a component of HOT, we propose investigations at the most detailed level. For practitioners looking for ways to foster HOT, CT, and CAT, it is important to know and hence take into account that definitions (and hence subsequently, measurements) are ambiguous. Given that the results seem positive in terms of the capability of P(j)BL to foster HOT and CT, it is important to guard that the child is not thrown away with the bathwater. Exposure to problems and collaboration with fellow students seem beneficial for fostering thinking skills.

Secondly, the lack of studies investigating CAT in P(j)BL means that this form of thinking is not yet embedded in the P(j)BL research literature. Given the characteristics of the P(j)BL process, CAT processes should be an object of investigation in these learning environments to further advance the theoretical understanding of CAT in P(j)BL.

Conclusion

In sum, the present review study led to several conclusions regarding conceptualizations of HOT, CT, and CAT in the P(j)BL literature. First, CAT is not embedded, and HOT is not frequently studied in the P(j)BL literature. Second, more focus lies on CT skills compared to CT dispositions in the research literature. Next, a lack of clear conceptualizations of HOT, CT, and CAT complicates the measurements and findings. This lack of conceptual clarity carries into instruments and tools of assessment that are limited in number (and sometimes availability) and of questionable validity. The lack of conceptual clarity also extends to P(j)BL environments, where essential components of PBL and PjBL were not always articulated or addressed in studies claiming to implement these approaches. Design issues in effect studies add to these complications. Further, the reference to HOT or CT skills conflicts with the literature on what differentiates skills from more intentional and purposefully implemented processes (i.e., cognitive and metacognitive strategies). Finally, mainly positive effects were found of P(j)BL on HOT and CT.

Appendix

Quantitative Analyses of CT Terms.

Table 5. CT Dispositions as Conceptualized in the Included Studies

Disposition	Terms used by studies included in this review	Count
CCTDI		
Truthseeking	active search of truth/ intellectual integrity/ prudence/ caution/ judicious/ likely to rethink and reconsider/ identification and exploration alternative frames of references/ objectivity/ willing to challenge information/ healthy skepticism	12
Open-mindedness	open mindedness/ sensitive to other ideas- understanding of own biases	7
Analyticity	being analytical	2
Systematicity	systematicity/ clarity on issues and ideas	3
Critical thinking self-confidence	confident in reasoning/ self-confidence/ tolerating ambiguity	3
Inquisitiveness	intellectual curiosity/ willing to investigate/ inquisitive/ inquisitive attitude/ valueing being well informed/ willing to seek rigorous results	6
Cognitive maturity	understanding of who they are/ maturity	3
Other		
Other (not CCTDI)	Self-regulated judgment/ metacognition, being knowledgeable, willing to suspend judgment, willingness to engage in sustained critical thinking, consistent internal motivation to value and use CT, cognitive dispositions, the will to evaluate a situation critically	8

CCTDI California Critical Thinking Disposition Inventory, *CT* Critical Thinking

Table 6. CT Skills as Conceptualized in the Included Studies

Skill	Terms used by studies included in this review	Count
CCTST		
Analysis	Analysis (of arguments)/ basic clarification/identification/ Obtain information/ pursue and comprehend relevant information/ interpretation	19
Evaluation	Evaluation/credibility assessment of statements made by others/ read and evaluate scientific info (appraise)/ judge the soundness of information/ conceptual, methodological, criteriological, contextual considerations upon which that judgment is based	16
Inference	Inference/ assess conclusions/argumentation	9
Deduction	Deduction/ reasoning	4
Induction	Induction	2
WGCTA		
Inference	Inference/ assess conclusions	8
Recognition of assumptions	Assumption identification/recognition of assumption/ noticing assumptions	3
Deduction	Deduction/ reasoning	4
Interpretation	Interpretation/analysis (of arguments)	15
Evaluation of arguments	Evaluation/ credibility assessment of statements made by others/ conceptual, methodological, criteriological, contextual considerations upon which that judgment is based/ argumentation	15
CCTT		
Induction – judging conclusions	Induction/Conclusion	5
Deduction	Deduction	3

Table 6. (continued)

Skill	Terms used by studies included in this review	Count
Credibility	Evaluation/credibility assessment of statements made by others/ read and evaluate scientific info (appraise)/ judge the soundness of information/ Conceptual, methodological, criteriological, contextual considerations upon which that judgment is based	16
Identification of assumptions	Assumption identification/recognition of assumption/ noticing assumptions	3
Semantics	Meaning	1
Other		Count
	(purposeful) self-regulatory judgment/ problem solving/ decision making/ creative thinking/ metacognition/ continuation clarification/ taking charge of the structures inherent in thinking and imposing intellectual standards upon them/ collect data/ observation/ use of higher-level thinking/ use of critical thinking processes/ use of tools for critical thinking/ analysis of psychological research/ ability to ask question/ answer questions	19

CCTST California Critical Thinking Skills Test, *WGCT7A* Watson–Glaser Critical Thinking Appraisal, *CCTT* Cornell Critical Thinking Test

Data Availability The data that support the findings of this study are available from the corresponding author, S.L., upon reasonable request.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest to disclose.

References

References preceded by an * were included in the review

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*(2), 275–314. https://doi.org/10.3102/0034654314551063

Alexander, P. A. (2014). Thinking critically and analytically about critical-analytic thinking. *Educational Psychology Review, 26*, 469–476. https://doi.org/10.1007/s10648-014-9283-1

Authors. (2022). [Blinded for review].

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. In L. Wilkerson & W. H. Gijselaers (Eds.), New Directions in Teaching and Learning: Issue 68. Bringing problem-based learning to higher education: Theory and practice (pp. 3–12). Jossey-Bass. https://doi.org/10.1002/tl.37219966804

Bataineh, R. F., & Zghoul, L. H. (2006). Jordanian TEFL graduate students' use of critical thinking skills (as measured by the Cornell Critical Thinking Test, Level Z). *International Journal of Bilingual Education and Bilingualism, 9*(1), 33–50. https://doi.org/10.1080/13670050608668629

Bezanilla, M. J., Galindo-Domínguez, H., & Poblete, M. (2021). Importance of teaching critical thinking in higher education and existing difficulties according to teacher's views. REMIE-Multidisciplinary Journal of Educational Research, 11(1), 20–48. https://doi.org/10.4471/remie.2021.6159

Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347–364. https://doi.org/10.1007/BF00138871

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. David McKay Co.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3–4), 369–398. https://doi.org/10.1080/00461520.1991.9653139

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

*Burris, S., & Garton, B. L. (2007). Effect of instructional strategy on critical thinking and content knowledge: Using problem-based learning in the secondary classroom. Journal of Agricultural Education, 48(1), 106–116. https://files.eric.ed.gov/fulltext/EJ840072.pdf

Butler, H. A., & Halpern, D. F. (2020). Critical thinking impacts our everyday lives. In R. J. Sternberg & D. F. Halpern (Eds.), Critical thinking in psychology (2nd ed., pp. 152–172). Cambridge University Press. https://doi.org/10.1017/9781108684354.008

Byrnes, J. P., & Dunbar, K. N. (2014). The nature and development of critical analytic thinking. *Educational Psychology Review, 26*, 477–493. https://doi.org/10.1007/s10648-014-9284-0

Caldwell, D. J., Caldwell, D. Y., Mcelroy, A. P., Manning, J. G., & Hargis, B. M. (1998). BASP-induced suppression of mitogenesis in chicken, rat and human PBL. *Developmental & Comparative Immunology, 22*(5–6), 613–620. https://doi.org/10.1016/S0145-305X(98)00037-8

*Carriger, M. S. (2016). What is the best way to develop new managers? Problem-based learning vs. lecture-based instruction. The International Journal of Management Education, 14(2), 92–101. https://doi.org/10.1016/j.ijme.2016.02.003

Castle, A. (2006). Assessment of the critical thinking skills of student radiographers. *Radiography, 12*(2), 88–95. https://doi.org/10.1016/j.radi.2005.03.004

Chakravorty, S. S., Hales, D. N., & Herbert, J. I. (2008). How problem-solving really works. *International Journal of Data Analysis Techniques and Strategies, 1*(1), 44–59. https://doi.org/10.1504/IJDATS.2008.020022

Chen, C., & Yang, Y. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review, 26*, 71–81. https://doi.org/10.1016/j.edurev.2018.11.001

Chiu, C.-F. (2020). Facilitating K-12 teachers in creating apps by visual programming and project-based learning. International Journal of Engineering Technology in Learning, 15(1), 103–118. https://www.learntechlib.org/p/217066/

*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

*Cortázar, C., Nussbaum, M., Harcah, J., Alvares, J., López, F., Goñi, J., & Cabezas, V. (2021). Promoting critical thinking in an online, project-based course. Computers in Human Behavior, 119, Article 106705. https://doi.org/10.1016/j.chb.2021.106705

*da Costa Carbogim, F., Barbosa, A. C. S., de Oliviera, L. B., De Sá Diaz, F. B. B., Toledo, L. V., Alves, K. R., de Castro Friedrich, D. B., Luiz, F. S., & de Araújo Püschel, V. A. (2018). Educational intervention to improve critical thinking for undergraduate nursing students: A randomized clinical trial. *Nurse Education in Practice, 33*, 121–126. https://doi.org/10.1016/j.nepr.2018.10.001

Dahlgren, M. A., & Dahlgren, L. O. (2002). Portraits of PBL: Students' experiences of the characteristics of problem-based learning in physiotherapy, computer engineering and psychology. *Instructional Science, 30*, 111–127. https://doi.org/10.1023/A:1014819418051

*Dakabesi, D., & Luoise, I. S. Y. (2019). The effectiveness of problem-based learning model to increase the students' critical thinking skills. Journal of Education and Learning, 13(4), 543–549. https://doi.org/10.11591/edulearn.v13i4.12940

David, J. L. (2008). Project-based learning. *Educational Leadership: Journal of the Department of Supervision and Curriculum Development, 65*(5), 80–82.

*Dilek Eren, C., & 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.

*Ding, X.-W. (2016). The effect of WeChat-assisted problem-based learning on the critical thinking disposition of EFL learners. *International Journal of Emerging Technology in Learning, 11*(12), 23–29. https://doi.org/10.3991/ijet.v11i12.5927

Dinsmore, D. L., & Fryer, L. (2022). *The dynamic interface between cognitive and metacognitive processing and higher-order, critical, or critical-analytic thinking [Manuscript Submitted for Publication]*. University of North Florida.

Dolmans, D., Loyens, S. M. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: A review of the literature. *Advances in Health Sciences Education, 21*(5), 1087–1112. https://doi.org/10.1007/s10459-015-9645-6

Duncan, T. G., & McKeachie, W. J. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist, 40*(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6

Duval, S. J., & Tweedie, R. L. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.2307/2669529

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Ennis, R. H. (1993). Critical thinking assessment. Theory Into Practice, 32(3), 179–186. https://www.jstor.org/stable/1476699

Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review, 32*, 81–111.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). W H Freeman/ Times Books/ Henry Holt & Co.

Ennis, R. (2011). Critical thinking: Reflection and perspective part I. *Inquiry: Critical Thinking Across the Disciplines, 26*(1), 4–18. https://doi.org/10.5840/inquiryctnews20112613

Facione, P. A. (1990a). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. California State University. https://eric.ed.gov/?id=ED315423

Facione, P. A. (1991). Using the California Critical Thinking Skills Test in research, evaluation, and assessment. California Academic Press. https://eric.ed.gov/?id=ED337498

Facione, P. A., Facione, N. C., & Giancarlo, C. A. (2000). 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. (1990b). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. The Delphi Report. California Academic Press.

*Fitriani, A., Zubaidah, S., Susilo, H., & Al Muhdhar, M. H. I. (2020). PBLPOE: A learning model to enhance students' critical thinking skills and scientific attitudes. International Journal of Instruction, 13(2), 89–106. https://doi.org/10.29333/iji.2020.1327a

*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

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27–61. https://doi.org/10.3102/00346543075001027

Glaser, E. M. (1941). An experiment in the development of critical thinking. *Teachers College Record, 43*(5), 1–18. https://doi.org/10.1177/016146814204300507

Grant, M. M., & Branch, R. M. (2005). Project-based learning in a middle school: Tracing abilities through the artifacts of learning. *Journal of Research on Technology in Education, 38*(1), 65–98. https://doi.org/10.1080/15391523.2005.10782450

Halpern, D. F. (2014). Thought and knowledge: An introduction to critical thinking (5th ed.). Psychology Press.

*Hassanpour Dehkordi, A., & 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://pjms.com.pk/issues/octdec108/article/article19.html

Helle, L., Tynjälä, P., & Olkinuora, E. (2006). Project-based learning in secondary education – theory, practice and rubber sling slots. *Higher Education, 51*, 287–314. https://doi.org/10.1007/s10734-004-6386-5

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*, 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hung, W., Jonassen, D. H., & Liu, R. (2008). Problem-based learning. In J. M. Spector, M. D. Merrill, J. Van Merriënboer, & M. P. Driscoll (Eds.), Handbook of research on educational communications and technology (pp. 485–506). https://doi.org/10.4324/9780203880869.ch38

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

Hung, W., Dolmans, D. H., & Van Merriënboer, J. J. (2019). A review to identify key perspectives in PBL meta-analyses and reviews: Trends, gaps and future research directions. *Advances in Health Sciences Education, 24*, 943–957. https://doi.org/10.1007/s10459-019-09945-x

*Iwaoka, W. T., Li, Y., & Rhee, W. Y. (2010). 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*(3), 68–75. https://doi.org/10.1111/j.1541-4329.2010.00100.x

Kivela, J., & Kivela, R. J. (2005). Student perceptions of an embedded problem-based learning instructional approach in a hospitality undergraduate programme. *International Journal of Hospitality Management, 24*, 437–464. https://doi.org/10.1016/j.ijhm.2004.09.007

Koh, G. C. H., Khoo, H. E., Wong, M. L., & Koh, D. (2008). The effects of problem-based learning during medical school on physician competency: A systematic review. *Canadian Medical Association Journal, 178*(1), 34–41. https://doi.org/10.1503/cmaj.070565

Krajcik, J. (2015). Project-based science. Science Teacher, 82(1), 25–27. http://www.nsta.org/store/product_detail.aspx?id=https://doi.org/10.2505/4/tst15_082_01_25

Krajcik, J., McNeill, K. L., & Reiser, B. J. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education, 92*(1), 1–32. https://doi.org/10.1002/sce.20240

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–264. https://doi.org/10.1207/s15430421tip4104_2

Larmer, J., Mergendoller, J., & Boss, S. (2015). Gold standard PBL: Essential project design elements. https://www.pblworks.org/blog/gold-standard-pbl-essential-project-design-elements

Leach, S. M., Immekus, J. C., French, B. F., & Hand, B. (2020). The factorial validity of the Cornell Critical Thinking Tests: A multi-analytical approach. *Thinking Skills and Creativity, 37*, 100676. https://doi.org/10.1016/j.tsc.2020.100676

Lloyd-Jones, G., & Hak, T. (2004). Self-directed learning and student pragmatism. *Advances in Health Sciences Education, 9*, 61–73. https://doi.org/10.1023/B:AHSE.0000012228.72071.1e

Lloyd-Jones, G., Margetson, D., & Bligh, J. G. (1998). Problem-based learning: A coat of many colours. *Medical Education, 32*, 492–494. https://doi.org/10.1046/j.1365-2923.1998.00248.x

Lombardi, D., Shipley, T. F., Astronomy Team, Biology Team, Chemistry, Engineering Team, Geography Team, Geoscience Team, & Physics Team. (2021). The curious construct of active learning. Psychological Science in the Public Interest, 22(1) 8–43. https://doi.org/10.1177/1529100620973974

Lombardi, D. (2022). On the horizon: The promis and power of higher-order, critical, and critical-analytical thinking. [Manuscript Submitted for Publication]. University of Maryland.

Lombardi, D., Matewos, M. M., Jaffe, J., Zohery, V., Mohan, S., Bock, K., & Jamani, S. (2022). *Discourse and agency during scaffolded middle school science instruction*. Advance online publication. https://doi.org/10.1080/0163853X.2022.2068317

Loyens, S. M. M., Kirschner, P. A., & Paas, F. (2012). Problem-based learning. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), APA educational psychology handbook, Vol. 3. Application to learning and teaching (pp. 403–425). American Psychological Association. https://doi.org/10.1037/13275-016

Loyens, S. M. M., & Rikers, R. M. J. P. (2017). Instruction based on inquiry. In R. E. Mayer & P. A. Alexander (Eds.), Handbook of research on learning and instruction (2nd ed., pp. 405–431). Routledge.

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review, 20*, 411–427. https://doi.org/10.1007/s10648-008-9082-7

*Lyons, E. M. (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*(1), 21. https://doi.org/10.2202/1548-923X.1524

Macpherson, K., & Owen, C. (2010). Assessment of critical thinking ability in medical students. *Assessment of Evaluation in Higher Education, 35*(1), 45–54. https://doi.org/10.1080/02602930802475471

Mathews, S. R., & Lowe, K. (2011). Classroom environments that foster a disposition for critical thinking. *Learning Environments Research, 14*(1), 59–73. https://doi.org/10.1007/s10984-011-9082-2

Maudsley, G. (1999). Do we all mean the same thing by "problem-based learning"? A review of the concepts and a formulation of the ground rules. *Academic Medicine, 74*(2), 178–185. https://doi.org/10.1097/00001888-199902000-00016

McNeill, K. L., & Krajcik, J. S. (2011). Supporting Grade 5–8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing. Pearson.

Miri, B., David, B. C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education, 37*, 353–369. https://doi.org/10.1007/s11165-006-9029-2

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*, 1. https://doi.org/10.1186/2046-4053-4-1

*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

Murphy, P. K., Rowe, M. L., Ramani, G., & Silverman, R. (2014). Promoting critical-analytic thinking in children and adolescents at home and in school. *Educational Psychology Review, 26*, 561–578. https://doi.org/10.1007/s10648-014-9281-3

Nagarajan, S., & Overton, T. (2019). Promoting systems thinking using project- and problem-based learning. *Journal of Chemical Education, 96*(12), 2901–2909. https://doi.org/10.1021/acs.jchemed.9b00358

Norman, G. R., & Schmidt, H. G. (2000). Effectiveness of problem-based learning curricula: Theory, practice and paper darts. *Medical Education, 34*(9), 721–728. https://doi.org/10.1046/j.1365-2923.2000.00749.x

Olivares, S., Saiz, C., & Rivas, S. F. (2013). Encouragement for thinking critically. *Electronic Journal of Research in Education, 11*(2), 367–394. https://doi.org/10.14204/ejrep.30.12168

Oliver, P., (2012). Succeeding with your literature review: A handbook for students. McGraw Hill.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan: A web and mobile app for systematic reviews. *Systematic Reviews, 5*, 210. https://doi.org/10.1186/s13643-016-0384-4

*Ozturk, C., Muslu, G. K., & Dicle, A. (2008). A comparison of problem-based and traditional education on nursing students' critical thinking disposition. *Nurse Education Today, 28*(5), 627–632. https://doi.org/10.1016/j.nedt.2007.10.001

*Pardamean, B. (2012). Measuring change in critical thinking skills of dental students educated in a PBL curriculum. *Journal of Dental Education, 76*(4), 443–453. https://doi.org/10.1002/j.0022-0337.2012.76.4.tb05276.x

Pellegrino, J. W., & Hilton, M. L. (Eds.) (2012). Education for life and work: Developing transferable knowledge and skills in the 21st Century. National Academies Press. http://www.nap.edu/catalog.php?record_id=13398

Razali, S. N., Noor, H. A. M., Ahmad, M. H., & Shahbodin, F. (2017). Enhanced student soft skills through integrated online project based collaborative learning. *International Journal of Advanced and Applied Sciences, 4*(3), 59–67. https://doi.org/10.21833/ijaas.2017.03.010

*Rehmat, A. P., & Hartley, K. (2020). Building engineering awareness: Problem-based learning approach for STEM integration. Interdisciplinary Journal of Problem-Based Learning, 14(1). https://doi.org/10.14434/ijpbl.v14i1.28636

Saad, A., & Zainudin, S. (2022). A review of project-based learning (PBL) and computational thinking (CT) in teaching and learning. *Learning and Motivation, 78*, 101802. https://doi.org/10.1016/j.lmot.2022.101802

*Santiprastikul, S., Sithivong, K., & Polnueangma, O. (2013). The first year nursing students' achievement and critical thinking in local wisdom course using problem based learning process. *Wireless Personal Communications, 69*, 1077–1085. https://doi.org/10.1007/s11277-013-1067-2

*Saputro, A. D., Atun, S., Wilujeng, I., Ariyanto, A., & Arifin, S. (2020). Enhancing pre-service elementary teachers' self-efficacy and critical thinking using problem-based learning. European Journal of Educational Research, 9(2), 765–773. https://doi.org/10.12973/eu-jer.9.2.765

*Sasson, I., Yehuda, I., & Malkinson, N. (2018). Fostering the skills of critical thinking and question-posing in a project-based learning environment. *Thinking Skills and Creativity, 29*, 203–212. https://doi.org/10.1016/j.tsc.2018.08.001

Sawyer, R. K. (Ed.). (2014). The Cambridge Handbook of the Learning Sciences. Cambridge University Press.

Schmidt, H. G. (2000). Assumptions underlying self-directed learning may be false. *Medical Education, 34*(4), 243–245. https://doi.org/10.1046/j.1365-2923.2000.0656a.x

Schmidt, H. G., Loyens, S. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and. *Educational Psychologist, 42*(2), 91–97. https://doi.org/10.1080/00461520701263350

Schmidt, H. G., Van der Molen, H. T., Te Winkel, W. W. R., & Wijnen, W. H. F. W. (2009). Constructivist, problem-based learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educational Psychologist, 44*(4), 227–249. https://doi.org/10.1080/00461520903213592

Schraw, G., McCrudden, M. T., Lehman, S., & Hoffman, B. (2011). An overview of thinking skills. In G. Schraw & D. H. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 19–45). Information Age Publishing.

Schraw, G., & Robinson, D. H. (2011). Conceptualizing and assessing higher order thinking skills. In G. Schraw & D. H. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 1–15). Information Age Publishing.

*Şendağ, S., & Odabaşı, 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

Servant-Miklos, V. F. C. (2019). Fifty years on: A retrospective on the world's first problem-based learning programme at McMaster University Medical School. *Health Professions Education, 5*(1), 3–12. https://doi.org/10.1016/j.hpe.2018.04.002

Si, J., Kong, H., & Lee, S. (2019). Developing clinical reasoning skills through argumentation with the concept map method in medical problem-based learning. *Interdisciplinary Journal of Problem-Based Learning, 13*(1), 5. https://doi.org/10.7771/1541-5015.1776

*Siew, N. M., & Mapeala, R. (2016). The effects of problem-based learning with thinking maps on fifth graders' science critical thinking. Journal of Baltic Science Education, 15(5), 602–616. https://www.scientiasocialis.lt/jbse/?q=node/527

Smits, P. B. A., Verbeek, J. H. A. M., & de Buisonjé, C. D. (2002). Problem based learning in continuing medical education: A review of controlled evaluation studies. *British Medical Journal, 324*, 153–156. https://doi.org/10.1136/bmj.324.7330.153

Spaulding, W. B. (1969). The undergraduate medical curriculum (1969 model): McMaster University. *Canadian Medical Association Journal, 100*, 659–664.

Strobel, J., & Van Barneveld, A. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary Journal of Problem-Based Learning, 3*, 44–58. https://doi.org/10.7771/1541-5015.1046

*Sugeng, B., & Suryani, A. W. (2020). Enhancing the learning performance of passive learners in a financial management class using problem-based learning. Journal of University Teaching & Learning Practice, 17(1), Article 5. https://doi.org/10.53761/1.17.1.5

Sungur, S., & Tekkaya, C. (2006). Effects of problem-based learning and traditional instruction on self-regulated learning. *The Journal of Educational Research, 99*(5), 307–317. https://doi.org/10.3200/JOER.99.5.307-320

Tal, T., Krajcik, J. S., & Blumenfeld, P. C. (2006). Urban schools' teachers enacting project-based science. *Journal of Research in Science Teaching, 43*(7), 722–745. https://doi.org/10.1002/tea.20102

*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), Article 769. https://hdl.handle.net/10520/EJC148686

Thomas, J. W. (2000). *A review of research on project-based learning*. http://www.bie.org/files/researchreviewPBL.pdf.

*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

Torff, B. (2011). Critical thinking in the classroom: Teachers' beliefs and practices in instruction and assessment. In G. Schraw & D. H. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 361–394). Information Age Publishing.

Van Peppen, L. M. (2020). Fostering critical thinking: Generative processing strategies to avoid bias in reasoning [Doctoral dissertation, Erasmus University Rotterdam]. RePub. hdl.handle.net/1765/130461

Watson, G., & Glaser, E. M. (1980). Watson-Glaser Critical Thinking Appraisal: Forms A and B manual. The Psychological Corporation.

Watson, G., & Glaser, E. M. (2009). Watson–Glaser^TM II Critical Thinking Appraisal: Technical manual and user's guide. Pearson.

Watson, G., & Glaser, E. M. (1994). *Watson-Glaser Critical Thinking Appraisal*. The Psychological Corporation.

Wijnia, L., Loyens, S. M. M., & Rikers, R. M. J. P. (2019). The problem-based learning process: An overview of different models. In M. Moallem, W. Hung, and N. Dabbagh (Eds.), The Wiley Handbook of problem-based learning (pp. 273–295). John Wiley & Sons. https://doi.org/10.1002/9781119173243.ch12

Yeh, M.-L. (2002). Assessing the reliability and validity of the Chinese version of the California Critical Thinking Disposition Inventory. *International Journal of Nursing Studies, 39*(2), 123–132. https://doi.org/10.1016/S0020-7489(01)00019-0

Yen, T. S., & Halili, S. H. (2015). Effective teaching of higher-order thinking (HOT) in education. The Online Journal of Distance Education and e-Learning, 3(2), 41–47. https://tojdel.net/journals/tojdel/articles/v03i02/v03i02-04.pdf

Yetkiner, Z. E., Anderoglu, H., & Capraro, R. M. (2008). *Research summary: Project-based learning in middle grades mathematics*. http://www.nmsa.org/Research/ResearchSummaries/ProjectBasedLearninginMath/tabid/1570/Default.aspx.

*Yu, W.-C. W., Lin, C. C., Ho, M.-H., & Wang, J. (2015). Technology facilitated by PBL pedagogy and its impact on nursing students' academic achievement and critical thinking dispositions. Turkish Online Journal of Educational Technology, 14(1), 97–107. https://eric.ed.gov/?id=EJ1057343

*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., Kunaviktikul, W., Klunklin, A., & Williams, B. A. (2008b). Promoting critical thinking skills through problem-based learning. Chiang Mai University Journal of Social Science and Humanities, 2(2), 85–99. https://www.thaiscience.info/journals/Article/CMUS/10613690.pdf

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

Zabit, M. N. M. (2010). Problem-based learning on students critical thinking skills in teaching business education in Malaysia: A literature review. *American Journal of Business Education, 3*(6), 19–32. https://doi.org/10.19030/ajbe.v3i6.436

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In Handbook of Metacognition in Education (pp. 299–315). Routledge.