

Effects of online peer assessment on higher-order thinking: A meta-analysis

Ying Zhan | Zi Yan | Zhi Hong Wan  | Xiang Wang |
Ye Zeng | Min Yang | Lan Yang 

Department of Curriculum and Instruction,
The Education University of Hong Kong,
Hong Kong SAR, China

Correspondence

Zhi Hong Wan, Department of Curriculum and Instruction, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, Hong Kong SAR, China.
Email: wanzh@eduhk.hk

Funding information

Research Cluster Fund, Grant/Award Number: RG76/2020-2021R

Abstract

Online peer assessment (OPA) has been increasingly adopted to develop students' higher-order thinking (HOT). However, there has not been a synthesis of research findings on its effects. To fill this gap, 17 papers (published from 2000 to 2022) that reported either a comparison between a group using OPA ($n = 7$; $k = 22$) and a control group or a pre–post comparison ($n = 10$; $k = 17$) were reviewed in this meta-analysis. The overall effect of OPA on HOT was significant ($g = 0.76$). Furthermore, OPA exerted more significant effects on convergent HOT (eg, critical thinking, reasoning and reflective thinking; $g = 0.97$) than on divergent HOT (eg, creativity and problem-solving; $g = 0.38$). Reciprocal roles and anonymity were found to positively moderate the impacts of OPA on HOT, although their moderating effects were not statistically significant because of small sample size of studies in the analysis. The results of the meta-analysis reinforce the arguments for regarding OPA as a powerful learning tool to facilitate students' HOT development and reveal important factors that should be considered when adopting OPA to enhance students' HOT.

KEYWORDS

convergent thinking, divergent thinking, higher-order thinking, meta-analysis, online peer assessment

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *British Journal of Educational Technology* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Practitioner notes

What is already known about this topic

- Online peer assessment (OPA) has significant positive impacts on learning achievement.
- OPA has been regarded as a potential approach to cultivating students' higher-order thinking (HOT) but has not been proved by meta-analysis.
- OPA should be carefully designed to maximise its effectiveness on learning.

What this paper adds

- OPA has been proved to significantly positively influence students' HOT via meta-analysis.
- OPA exerted more significant effects on convergent HOT than on divergent HOT.
- The potential of reciprocal roles and anonymity for moderating the impacts of OPA on HOT should not be underestimated.

Implications for practice and/or policy

- OPA could be a wise choice for practitioners when they help students to achieve a balanced development of HOT dispositions and skills.
- Students' divergent HOT can be encouraged in their uptake of peer feedback and by allowing them autonomy in deciding assessment criteria.
- OPA with design elements of reciprocal roles and anonymity has great potential to promote students' HOT.

INTRODUCTION

Cultivating students' higher-order thinking (HOT) has been acknowledged as a significant education aim in the 21st century (Brookhart, 2010; Huang et al., 2022; Lin et al., 2021). Building up students' knowledge capacity may not satisfy the requirements of our ever-evolving and increasingly challenging society, fostering their HOT (eg, creative thinking, critical thinking, metacognition and problem-solving) is necessary in the new era (Miri et al., 2007; Newmann, 1991; Yen & Halili, 2015). HOT is beneficial for persons' successful learning, life and work as well as for the social development (Prayoonsri et al., 2015; Tanujaya et al., 2017). However, enhancing students' HOT is a challenging task for teachers (Miri et al., 2007). Teachers need to change their traditional textbook-based teaching and switch to students' active learning to provide chances of practising HOT and constructing their deep thoughts through social interactions (Liu et al., 2021; Richmond & Hagan, 2011; Yen & Halili, 2015).

Recently, online peer assessment (OPA) has been regarded as a potentially powerful approach to develop students' HOT. OPA requires students' active roles in assessment thus activating their multiple complex cognitive processes beyond the memorization of facts (Lin et al., 2021; van Popta et al., 2017). Meanwhile, social interactions between peers in the OPA process enable students to reconstruct their thinking and resolve cognitive conflicts (Anderson et al., 2001; Bruning et al., 2004; Guiller et al., 2008; Lu et al., 2022; Zhan, 2021).

However, no research has attempted to synthesise the effect size of OPA on HOT through meta-analysis. Furthermore, a significant gap remains in revealing how OPA interventions can be designed to maximise HOT development. It is necessary to draw together quantitative evidence about which commonly used design elements of OPA influence HOT development.

To respond to the above-mentioned needs, we statistically analyse the overall effect size of OPA and consider potential moderators in this meta-analysis paper.

LITERATURE REVIEW

Online peer assessment

Peer assessment is a two-way process in which students judge the work of their peers against assessment criteria, give constructive feedback as assessors, and in turn receive their peers' comment, critically evaluate its' quality, and refine their work accordingly as assessees (Cheng et al., 2015; Li et al., 2012). It actively engages students in social interaction and provides them a unique learning experience to learn better and prepare well for future life and work (Patchan et al., 2018; Pryor & Crossouard, 2008; van Gennip et al., 2009; van Popta et al., 2017). With the fast-moving information and communications technology revolution, OPA has become popular and is regarded as more powerful tool than face-to-face or paper-based one due to its technological affordances. With the aid of technology, students can engage in peer assessment at their own speed and convenience (Ciftci & Kocoglu, 2012; Shang, 2017). In addition, they can critically evaluate each other's work in an anonymous way (Lin, 2018; Raes et al., 2015) and have asynchronous or synchronous communication about the received comments (Chen, 2016; Gikandi & Morrow, 2016).

OPA has great potential to promote students' learning in many various ways; thus, it has educational values for teachers (Topping, 1998; van Gennip et al., 2009; Zhan et al., 2022). It can improve students' academic performance in terms of subject-based knowledge and skills (van Zundert et al., 2010), HOT such as critical thinking and metacognition (van Popta et al., 2017), social-affective development such as learning motivation and attitudes (Noroozi & Mulder, 2017) and feedback literacy (Ketonen et al., 2020).

There have been some research syntheses of the effects of OPA on learning. Tenório et al. (2016) found 60% of the papers they reviewed reported OPA-positive impacts on learning. Zhan et al. (2022) found a higher ratio of positive impacts of OPA on learning in their review of the studies in Chinese contexts. van Popta et al. (2017) focused on reviewing feedback providers' perspectives of the benefits of OPA and found that providing feedback online could help students develop their high-level cognition and metacognition. Zheng et al. (2020) meta-analysed 37 empirical studies and reported that OPA exerted a medium-sized positive effect on students' learning achievement. Li et al. (2020) compared the effects of OPA with those of paper-based peer assessment in their meta-analysis and found that OPA generated greater learning gains.

The above-mentioned literature review studies focus on synthesising the effects of OPA on learning achievement majorly concerning domain-specific knowledge and learning skills. Although van Popta et al. (2017) and Zhan et al. (2022) have claimed learning benefits of OPA on HOT, the synthesis of such effects is kind of rough judgement relying on the authors' reported results instead of meta-analysis. A synthesis of the findings on the effects of OPA on HOT is called for to advance relevant knowledge.

OPA and HOT

HOT is the cognitive capacity to go beyond the given information to classify, infer, generalise and resolve problems in complicated situations (Bruner, 1973; Lewis & Smith, 1993; Resnick, 1987). It is an umbrella term for various forms of thinking including critical thinking, reflective thinking, argumentation, reasoning, problem-solving, creativity and metacognitive

thinking (Miri et al., 2007). In spite of its different forms, HOT has key features being indirect, non-algorithmic and complex (Lewis & Smith, 1993; Yen & Halili, 2015). Because of the common features shared by various forms of HOT, it is possible to merge them as an investigated object in this meta-analysis study. However, the subtle differences in different forms of HOT are worthwhile for in-depth investigation. Therefore, HOT can be further categorised according to thinking orientation into two groups namely convergent HOT and divergent HOT. Convergent HOT (eg, critical thinking, reflective thinking, argumentation, reasoning and metacognitive thinking) is mainly driven by logical reasoning and established multiple criteria while divergent HOT (eg, creativity and problem-solving) is typically more spontaneous, free-flowing and non-linear (Baker et al., 2001; Beyer, 1987; Runco, 2014).

Cultivating students' HOT is a difficult task for teachers and scholars have suggested a number of approaches to do it (Liu et al., 2021; Miri et al., 2007; Yen & Halili, 2015). Among these approaches, OPA has caught increasing attention from researchers and practitioners. OPA activities can activate multiple complex cognitive processes which are related to HOT. van Popta et al. (2017, p. 29) summarised the cognitive processes of providing peer feedback in which 'students compare and question ideas, evaluate, suggest modifications; and reflect, plan and regulate their own thinking'. They claimed that students developed HOT through practising OPA. Furthermore, peer assessment is a social activity that inevitably involves peer interaction and OPA facilitate such interaction (Zhan, 2021). Peer interaction brings about cognitive reconstruction or conflict solution thus causing students' HOT development (Anderson et al., 2001; Bruning et al., 2004; Guillier et al., 2008; Lu et al., 2022). Shin et al. (2020) demonstrated that peer interaction could deepen students' understanding of complex problems and create advanced solutions through meaningful interactions. Csikszentmihalyi (1996, p. 23) claimed 'creativity doesn't happen inside people's head but in the interaction between an individual's thoughts and a sociocultural context'. Zhan (2021) also argued that peer interaction in OPA might cause cognitive conflicts which is the key condition for critical thinking development.

To determine the effects of OPA on HOT, the measurement of HOT is important. The effects of OPA on HOT can be measured by examining the changes in thinking skills and dispositions since a 'good thinker' should be equipped with both cognitive ability and a tendency towards thoughtfulness (Newmann, 1991; Tishman et al., 1993). For example, critical thinking is often regarded as a construct consisting of both a dispositional and a cognitive dimension (Cheng & Wan, 2017; Sosu, 2013; West et al., 2008). In recent years, several empirical studies have been published to examine the extent to which OPA can develop students' HOT. Therefore, it is worthwhile doing a systematic synthesis of current empirical studies' findings to reinforce the argument on the power of OPA over HOT. Given thinking skills and dispositions are two different dimensions, it may be meaningful to synthesise the findings of existing research to examine whether the impacts of OPA differ for these two dimensions of HOT. Furthermore, as mentioned previously, HOT could be either convergent or divergent. It is also interesting to identify whether the impacts of OPA on convergent HOT and divergent HOT vary using meta-analysis.

Factors influencing the effects of OPA on students' HOT

The effects of OPA are different when it is designed and implemented differently (Li et al., 2020; Tenório et al., 2016; van Popta et al., 2017). Therefore, it is crucial to investigate the design elements that may moderate the effects of OPA interventions (Li et al., 2020; Panadero et al., 2018; Zhan et al., 2022). Many design elements of OPA have been stated in the literature. In this meta-analysis, we explore six design elements that have the potential to support HOT practice and enhance the peer interaction underpinning HOT development.

These elements are OPA mode, provision of assessment criteria, feedback content requirement, reciprocal roles, anonymity and dialogic interaction.

OPA could be conducted in different modes. Research has indicated that qualitative feedback (ie, comments only) is more conducive to learning than quantitative feedback (ie, grades or scores) (Brookhart, 2018). Reciprocal and simultaneous roles as assessor and assessee in the OPA process double students' opportunities to practise HOT (Tsivitanidou & Constantinou, 2016). Providing assessment criteria guides students to evaluate each other's work in a shared discourse, which leads to a comprehensive understanding of the work and in-depth discussion. Dominguez et al. (2015) found that participants valued assessment criteria provided on the OPA platform to facilitate their critical thinking practice. Feedback content has been viewed as another crucial factor influencing the effect of OPA on student learning (Cho & MacArthur, 2010). Filius et al. (2018) found in their study that the requirements of feedback content with explanation and suggestions made students think, reflect and review their answers.

Moreover, the design elements of anonymity and dialogic interaction influence the extent to which students engage in the OPA process. Anonymity is usually regarded as a good strategy for combating 'face culture' or friendship which inhibits peer interaction in peer assessment (Panadero & Alqassab, 2019; Zhan, 2021). Panadero and Alqassab (2019) found in their review that anonymity led to more critical peer feedback. Dialogic interaction in either an asynchronous or a synchronous mode involves students questioning, negotiating and decision-making about the peer feedback they receive (Zhan et al., 2022).

In addition to OPA design elements, the contexts where OPA takes place such as educational level and course learning environment could be related to the effects of peer assessment (Li et al., 2020). The methodological characteristics of studies such as experimental design, sample size and measurement tools may also influence the effects of OPA interventions. In this meta-analysis, we examined such moderators too.

RESEARCH QUESTIONS

The purposes of this meta-analytic review were to examine the overall effect of OPA, compare the effect of OPA on HOT dispositions with its effect on skills, and differentiate its effect on convergent HOT from divergent HOT. We also investigated whether the overall effect of OPA is moderated by OPA design elements. The research questions were as follows.

- RQ1: What is the overall effect of OPA on students' HOT?
- RQ2: What factors moderate the effect size of OPA on students' HOT?

METHODS

Searching studies

To identify relevant and reliable studies for this meta-analysis, EBSCO (including the Art & Architecture Source, British Education Index, Education Full Text, Education Research Complete, Eric, APA PsycArticles, APA PsycInfo, CINAHL with Full Text databases) and Web of science were searched. The time range of the search was January 2000 to February 2022. We did not find any relevant studies before 2000 in our pilot search.

To identify studies of HOT, we searched the abstracts of papers using the searching string "AB ('peer feedback' or 'peer comment' or 'peer review' or 'peer assessment' or 'peer editing' or 'peer revision' or 'peer correction' or 'peer evaluation' or 'peer marking' or 'peer

grading' or 'peer scoring' or 'peer rating') AND AB ('critical thinking' or 'higher-order thinking' or 'reasoning' or 'rational thinking' or 'problem solving' or 'reflective thinking' or 'reflection' or 'argumentation' or 'argumentative thinking' or 'metacognitive thinking' or 'creativity')". Altogether, 719 full-text English papers which were published in peer-reviewed journals were selected after cleaning duplicates from different data bases. We made the last search on 21 February 2022.

Selecting studies

The PRISMA flow chart (Moher et al., 2009) represents the process through which eligible articles were selected (see Figure 1). There were four inclusion criteria. First, the target effect of peer assessment should be related to HOT. A total of 675 papers were ruled out because they did not satisfy this requirement. Second, the studies should include an experimental design (ie, experimental/quasi-experimental design with control groups and repeated measures design without control groups). In all, 11 papers were ruled out according to this criterion. Third, the selected experimental studies should present enough data to calculate effect size. Nine papers were excluded because they did not provide sufficient data. Fourth,

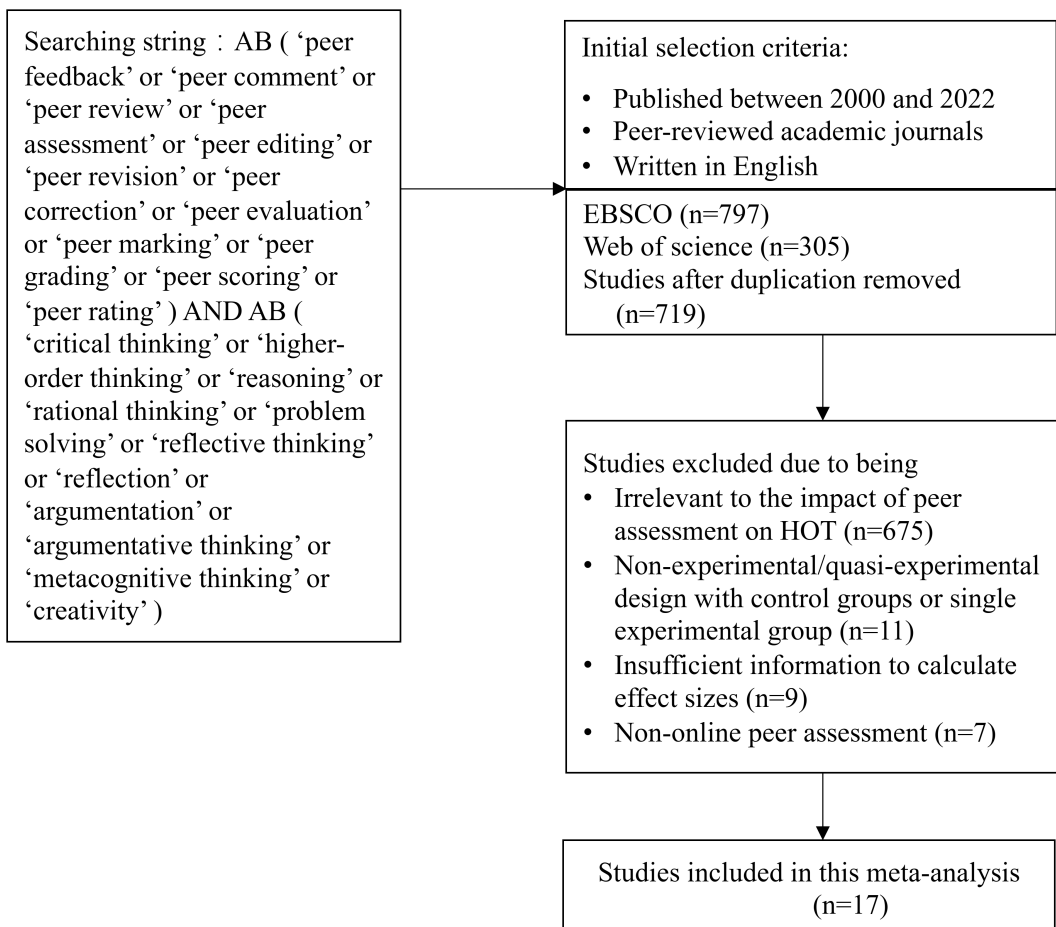


FIGURE 1 PRISMA flow chart displaying the process of paper selection and screening.

peer assessment should be conducted online. Consequently, seven papers were removed. Finally, 17 eligible articles were retained in the subsequent meta-analysis (Appendices A and B provide the descriptive characteristics of the studies and references).

Overview of the selected studies

The demographic information of the 17 selected papers is summarised in Appendix A. There is an increasing trend of experimental studies on the effects of OPA on HOT since 2012 and the number of relevant studies is the largest in 2021, with six articles. It is interesting to find that most of the identified studies were done in Asian countries or regions and more than half of the studies were conducted in Taiwan. 55.8% of the studies took place at tertiary level. OPA was applied either to a social science course or a science course. Furthermore, except one study did not mention the platform, half of the selected studies used self-developed OPF platforms (eg, a mobile learning system or an online interactive peer-review system) to enhance students' HOT. Others employed learning management systems (eg, Moodle and Cloud Classroom), social media (eg, blogs), Web-based platforms (eg, Peergrade and wikis) and sharing platforms (eg, Google Drive).

In all, 15 papers explored the effect of OPA on students' convergent thinking, including critical thinking (nine articles), argumentation (four articles), metacognitive thinking (two articles), reflection (one article) and reasoning (two articles). In comparison, six articles focused on divergent thinking, that is creativity (four articles) and problem-solving (two articles). Except for two articles that examined the effect of OPF on both HOT skills and disposition, the other 15 articles focus on either skills (10 articles) or dispositions (5 articles).

Coding procedure

The data regarding sample sizes, means and standard deviations were drawn out from each study to calculate effect size. A list of moderators was coded for each study to explicate the variation in the effect sizes. In all, 39 data sets were extracted from 18 studies to calculate effect size. This means that one study might contribute more than one effect size (Graham et al., 2015). When there were multiple control or experimental conditions in one study, we extracted multiple data sets based on the conditions. In addition, when one study explored more than one type of HOT, we extracted multiple data sets according to specific HOT types. For studies with an analytic rubric to assess one type of HOT, such as separate scores allocated to specific aspects of HOT without total scores, we extracted multiple data sets according to these aspects.

Four types of moderators were coded for each study. The first type of moderators was the impact target, which encompassed two variables (ie, thinking dimension being divided into thinking disposition and skills and thinking orientation being classified into convergent and divergent thinking) as discussed in the previous literature review. The second type of moderators was OPA design elements (see Table 1). The third type of moderator was regarding the contexts where the study was conducted including educational level (i.e. higher education and K–12 school education) and course type (ie, social science and science courses). The fourth type was the methodological characteristics of selected papers, including experimental design (ie, pretest–posttest design with control group, posttest-only design with control group and repeated measures design without control groups), sample size (large (≥ 100), medium ($\geq 50 < 100$) and small (< 50)) and instrument source (student work or questionnaire).

To secure coding reliability, each study was coded by Author 1 and Author 5. They independently coded every paper and then checked each other's coding. If there was a

TABLE 1 Coding scheme of OPA design elements as moderators.

OPA design elements	Code explanation
OPA mode	Grading a peer's work, giving comments on a peer's task performance or provision of both grade and comments
Provision of assessment criteria	Criteria such as rubrics, guidelines, a rating scale or checklist, etc. are provided on line when students do OPA
Feedback content requirement	Students need to provide specific comments such as strengths, weaknesses or suggestions
Reciprocal roles	Students play the dual roles of assessors and assessees
Anonymity	Students do not know who is assessing them or who they are assessing
Dialogical interaction	Students are involved in a continuous process of constructive dialogue about the feedback

discrepancy in coding, the two coders discussed the case till a consensus was achieved. Interrater agreement between the two coders was Cohen's Kappa = 0.89.

Statistical analysis

In the first stage, Cohen's d values (the standardised mean difference; Cohen, 1988) were calculated for all 39 sets of data identified from 17 studies to indicate their effect sizes. There were three different conditions for the calculation.

For the 10 data sets that included the post-test results of the experimental and control groups, the following formula was used:

$$d_1 = (\bar{x}_1 - \bar{x}_2) / S_{\text{pooled}} \quad (1)$$

$$S_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}} \quad (2)$$

$$SE_1 = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}} \quad (3)$$

where \bar{x}_1 is the reported mean outcome of the post-test for the experimental group; \bar{x}_2 is the reported mean outcome of the post-test for the control group; n_1 and n_2 are the sample sizes for the two groups; and s_1 and s_2 are the standard deviations for the groups.

Seven data sets included both pre- and post-test results for the experimental and control groups. In this case, change scores were used to calculate the effect sizes and \bar{x}_1 and \bar{x}_2 were obtained by subtracting the post mean from the pre mean. The standard deviation (SD) of the change (ie, s_1 and s_2) was transformed by the following formula, in which $Corr$ was assumed to be 0.5 between baseline and follow-up (Higgins et al., 2019):

$$SE_{\text{Change}} = \sqrt{SD_{E,\text{baseline}}^2 + SD_{E,\text{final}}^2 - (2 \times Corr \times SD_{E,\text{baseline}} \times SD_{E,\text{final}})} \quad (4)$$

In all, 22 data sets did not include control groups or missed data in the control groups (ie, adopted repeated measures design). In this case, effect sizes were calculated using the following formulae (Becker, 1988):

$$d_2 = \frac{\bar{x}_{t1} - \bar{x}_{t2}}{\sqrt{\frac{(s_{t1}^2 + s_{t2}^2)}{2}}}}{\sqrt{2(1 - r_{t1t2})}} \quad (5)$$

$$SE_2 = \sqrt{\frac{2(1 - r_{t1t2})}{n} + \frac{d^2}{2n}} \quad (6)$$

where \bar{x}_{t1} and \bar{x}_{t2} are the post-test and pre-test mean outcomes for the experimental group, respectively; s_{t1} and s_{t2} are the standard deviations for this group; n is the sample size for this group; and r_{t1t2} is the correlation between the outcomes at both time points. However, pre–post correlations of variables are rarely reported in published research, which is a drawback of this formula. Therefore, we calculated the effect size by imputing a correlation of 0.50.

It should be noted that a small sample size can cause an upward bias in Cohen's d effect. To correct the overestimation, Cohen's d values in this study were further transformed into Hedges' g through the following formula (Hedges, 1981):

$$G = d \times \left(1 - \frac{3}{4n - 9}\right) \quad (7)$$

Next, efforts were made to examine whether there were outliers, to provide evidence of the robustness of the findings. Normally, effect sizes that are beyond the range ($\bar{x} - 3SD$, $\bar{x} + 3SD$) are considered outliers (Acuna & Rodriguez, 2004). In this study, no outliers were identified.

As cautioned by Hedges and Vevea (1998), the true effect size may vary from study to study so a random-effect model was adopted to synthesise the 39 effect sizes generated in the preceding steps. In addition to calculating the overall effect size of the impacts of OPA on HOT, we explored the factors moderating the effect size. These moderators can be broadly classified into four types, including impact target, OPA design elements, the context of study and methodological characteristics. Meta-regression was conducted to calculate the moderation effect.

All analyses were performed in R. The effect size was calculated by the *dmetar* (Harrer et al., 2021) and *tidyverse* (Wickham et al., 2019) packages, the main meta-analysis was conducted using the *meta* package (Schwarzer, 2007), and meta-regression was examined using the *metafor* package (Viechtbauer, 2010).

RESULTS

Overall effect size

In all, 39 effect sizes within 17 studies reported a comparison between a control group and an experimental group ($n = 10$, $k = 17$) or a pre–post comparison of OPA ($n = 7$, $k = 22$). From the scatter plot of effect sizes in Figure 2, it can be seen that number of positive effect sizes exceeded that of negative ones. No outliers were detected in this group of studies (Hedges' g was above 3.13 or below -1.56). The overall effect of OPA intervention on HOT was 0.76, statistically different from zero (95%CI, 0.51–1.01; $t = 6.22$; $p < 0.01$, $k = 39$ in 17 studies).

Factors moderating the effect sizes of OPA on students' HOT

We analysed the factors moderating the effect sizes of OPA on HOT. The results of the meta-regression analysis are presented in Table 2. There is a slight difference between the impact on HOT disposition ($g = 0.83$, $p < 0.05$) than on HOT skills ($g = 0.72$, $p < 0.01$). Moreover, a statistically significant moderation effect was revealed for thinking orientation ($t = 5.88$, $p < 0.05$). It was found that the impact of OPA on convergent thinking ($g = 0.97$, $p < 0.01$) was

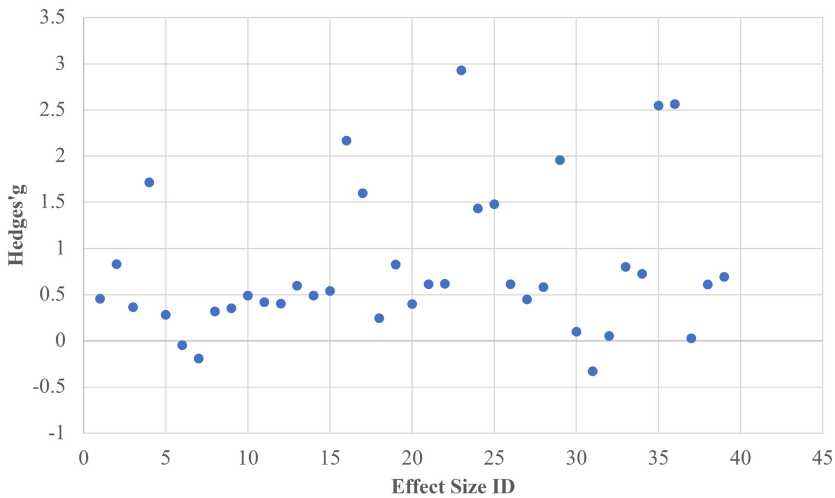


FIGURE 2 Scatter plot of effect sizes.

greater than the impact on divergent thinking ($g = 0.38$, $p < 0.05$), explaining 15.49% of the total heterogeneity.

Although statistically significant moderation effects of OPA design elements were not found in this study, considerable difference could be still observed for two design elements of OPA, including reciprocal roles and anonymity. The effect size was considerably bigger when students both received and provided feedback ($g = 0.83$, $p < 0.05$) than when they only received it ($g = 0.11$, $p < 0.01$). With respect to anonymity, the overall effect size of anonymous OPA ($g = 0.91$, $p < 0.05$) was higher than that in studies that did not adopt anonymous OPA ($g = 0.59$, $p < 0.05$).

The moderation effects of the other four OPA design elements were a relatively smaller. When assessment criteria were given, the effect size ($g = 0.78$, $p < 0.05$) was bigger than the effect size when students did not receive assessment criteria ($g = 0.44$, $p > 0.05$). The effect size was just slightly higher when both comments and grade were given ($g = 0.87$, $p < 0.01$), than when only comments ($g = 0.65$, $p < 0.05$) or grade was provided ($g = 0.74$, $p < 0.05$). Nonetheless, the effect size of studies with requirements for feedback content ($g = 0.79$, $p < 0.05$) was like those without such requirements ($g = 0.73$, $p < 0.01$). The effect size of OPA that featured interacting with feedback givers ($g = 0.79$, $p < 0.05$) was also similar to that of OPA which did not ($g = 0.63$, $p < 0.05$).

The effect size was larger for intervention in K–12 settings ($g = 0.98$, $p < 0.01$) than those in higher education settings ($g = 0.65$, $p < 0.05$). In terms of subject area, the g value for social science courses was 0.80 ($p < 0.01$), and the value for science courses was 0.73 ($p < 0.05$). For methodological characteristics of each study, the effect size of studies with pretest–posttest control group design ($g = 0.36$, $p > 0.05$) was smaller than the value of those with posttest-only control group design ($g = 0.74$, $p < 0.05$) which was similar to the studies with pretest–posttest design ($g = 0.90$, $p < 0.05$). The effect size showed a decline trend with the rise of sample size, small ($g = 0.92$, $p < 0.05$), medium ($g = 0.65$, $p < 0.05$) and large ($g = 0.55$, $p < 0.05$). Furthermore, the effect size appeared to be a bit smaller when data were collected by assignment ($g = 0.68$, $p < 0.01$) than questionnaire ($g = 0.84$, $p < 0.05$).

TABLE 2 Differences in effect sizes of OPA on HOT for moderators.

Moderators		No. of effect size	Estimate (95% CI)	Test statistic	p-value
<i>Impact target</i>					
Dimension of thinking	Skills	26	0.72 (0.19, 1.25)	$t_{(1,37)} = 0.17$	0.68
	Disposition	13	0.83 (0.40, 1.26)		
Thinking orientation	Convergent	26	0.95 (0.67, 1.24)	$t_{(1,37)} = 5.88$	0.02*
	Divergent	13	0.37 (-0.12, 0.86)		
<i>OPA design elements</i>					
OPA mode	Comment	17	0.65 (0.12, 1.19)	$t_{(2,36)} = 0.33$	0.72
	Grade	5	0.74 (-0.07, 1.54)		
	Both	17	0.87 (0.49, 1.25)		
Provision of assessment criteria	Without	3	0.44 (-0.46, 1.33)	$t_{(1,35)} = 0.63$	0.43
	With	34	0.78 (-1.30, 1.74)		
Feedback content requirement	Without	18	0.73 (0.35, 1.11)	$t_{(1,30)} = 0.04$	0.84
	With	14	0.79 (0.21, 1.36)		
Reciprocal roles	Without	4	0.11 (-0.64, 0.85)	$t_{(1,37)} = 3.50$	0.07
	With	35	0.83 (0.04, 1.62)		
Anonymity	Without	15	0.59 (0.22, 0.96)	$t_{(1,27)} = 1.47$	0.24
	With	14	0.91 (0.37, 1.45)		
Dialogical interaction	Without	21	0.63 (0.31, 0.95)	$t_{(1,32)} = 0.39$	0.54
	With	13	0.79 (0.27, 1.31)		
<i>Context of study</i>					
Educational level	K–12	13	0.98 (0.55, 1.40)	$t_{(1,37)} = 1.62$	0.21
	Higher education	26	0.65 (0.13, 1.17)		
Course	Social science	18	0.77 (0.40, 1.15)	$t_{(1,37)} = 0.01$	0.93
	Science	21	0.75 (0.25, 1.25)		
<i>Research design</i>					
Research design	Control group design	7	0.36 (-0.22, 0.93)	$t_{(2,36)} = 1.36$	0.27
	Repeated measures design	10	0.74 (0.23, 1.56)		
	Repeated & control design	22	0.90 (-0.02, 1.49)		
Sample size	Small	18	0.92 (0.15, 1.69)	$t_{(2,36)} = 0.74$	0.48
	Medium	16	0.65 (-0.13, 1.43)		
	large	5	0.55 (-0.12, 1.23)		
Data source	Assignment	20	0.68 (0.33, 1.03)	$t_{(1,37)} = 0.42$	0.52
	Questionnaire	19	0.84 (0.34, 1.34)		

Comparison among the effect sizes of different types of convergent and divergent HOT

Since significant difference was revealed for thinking orientation mentioned in the previous section, further comparison was made among different types of convergent HOT. As showed in Figure 3, the effect sizes of OPA on the two types of convergent HOT whose judgement

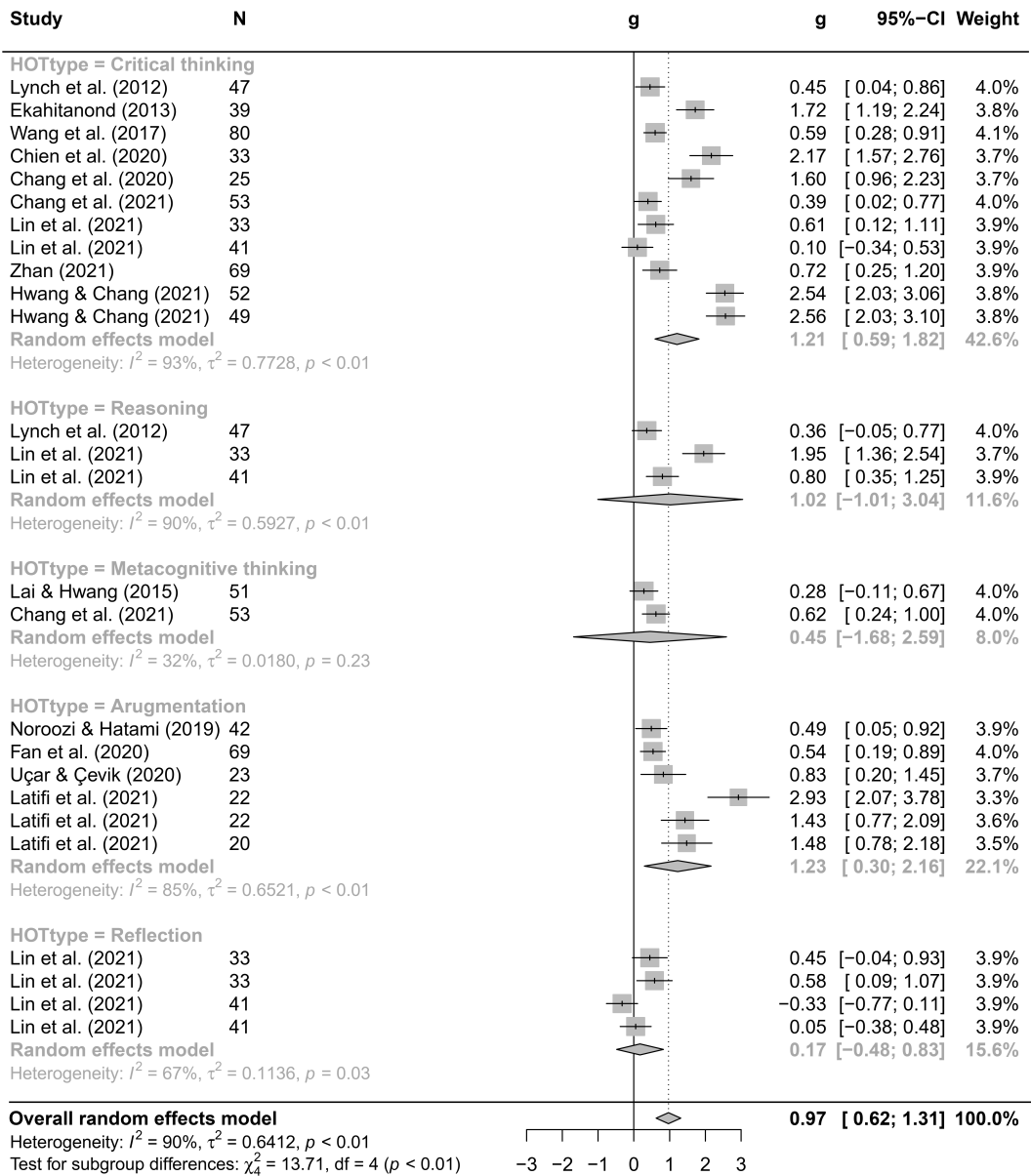


FIGURE 3 Forest plot of convergent thinking.

targets the thinkers themselves (ie, metacognitive thinking and reflection) looked smaller than the three types targeting other learners (ie, critical thinking, reasoning and argumentation). The g values of metacognitive thinking and reflection were, respectively, 0.45 ($p > 0.05$) and 0.17 ($p < 0.05$) while those values of critical thinking, reasoning and argumentation were 1.21 ($p < 0.01$), 1.02 ($p < 0.01$) and 1.23 ($p < 0.01$). On the contrary, Figure 4 indicated that the effect size of creativity ($g = 0.47$, $p > 0.05$) was rather close to that of problem-solving ($g = 0.23$, $p > 0.05$).

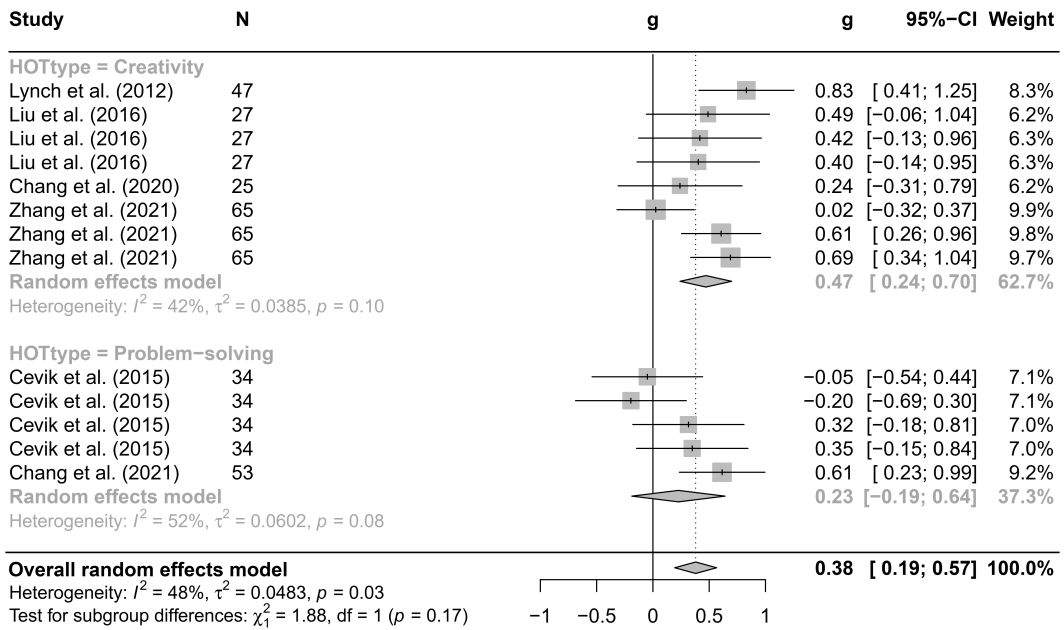


FIGURE 4 Forest plot of divergent thinking.

DISCUSSION AND CONCLUSION

This meta-analysis yields the first synthesis on the effects of OPA on HOT. Overall, the mean effect size of OPA interventions on HOT was 0.76, indicating a significant large effect size for this educational intervention (Hattie, 2009). This finding echoes previous meta-analysis of the impacts of OPA on learning achievement (Li et al., 2020; Zheng et al., 2020) and suggests that OPA deserves more pedagogical attention when educators and teachers contemplate effective ways of developing students' HOT.

The result indicated that OPA significantly increased students' HOT dispositions and skills but there was no significant difference between these two specific effects. Cheng and Wan (2017) argued that different approaches may have differentiated impacts on HOT dispositions and skill. In other words, an approach might help to promote HOT disposition but may not exert impacts on HOT skills. According to the finding of this meta-analysis, OPA seems an exceptional approach and could be a wise choice for practitioners when they help students to achieve a balanced development of HOT.

More interestingly, the meta-analysis revealed that the effect size of OPA on convergent HOT was significantly larger than that on divergent HOT. There are two possible explanations for this result. First, OPA requires more convergent HOT processes than divergent HOT processes. Li and Grion (2019) analysed the cognitive processes of giving and receiving peer feedback. They found that when giving peers feedback, students were involved in 'comparison, feedback co-construction, and reflective thinking processes' and when receiving peer feedback, they were engaged in 'feedback evaluation, reflection, self-assessment, and awareness of effective feedback' (Li & Grion, 2019, p. 13). Those findings reveal that OPA involves convergent HOT processes of judging, reflection, critical thinking and self-regulation. Divergent HOT, such as creative thinking and problem-solving, might be activated when students' construct suggestions about their peers' work and follow-up on peer feedback (Zhang et al., 2021). Second, students often evaluate each other's work against certain assessment criteria, which leads to a shared discourse to judge the quality of work and reflect on how to improve

(Orsmond et al., 2000; Topping, 2003; van den Berg et al., 2006). In this meta-analysis, 82% of the reviewed papers ($n = 14$) reported the provision of assessment criteria in OPA activities. However, assessment criteria might constrain students' divergent thinking, such as creativity (Cockett & Jackson, 2018; Gezie et al., 2012; Leader & Clinton, 2018). Our findings have at least two implications for practitioners who want to promote students' divergent HOT via OPA. First, students' divergent HOT can be encouraged in their uptake of feedback. Practitioners can enhance this phase by supporting and monitoring students' uptake of peer feedback. Second, practitioners can use OPA without criteria or allow students the autonomy to express ideas or suggestions that are not required by assessment criteria.

We further scrutinised the subtle differences among different types of convergent HOT. It is interesting to observe from forest plot of convergent HOT that OPA seems more conducive to critical thinking, reasoning and argumentation which targets other learners than to metacognitive thinking and reflective thinking which is self-targeting. In fact, when students are engaged in the OPA activities, they have to make evaluative judgement of each other's work during which they integrate their multiple strands of knowledge in an analytical way, make decisions and justify their decisions (Tai et al., 2018). Therefore, HOT such as reasoning, argumentation and critical thinking are inevitably practised during such a decision-making process. Although evaluative judgement can also lead to metacognitive and reflective thinking, such activation could be effectively triggered by self-assessment (Cowan, 2010). Therefore, a blended method of peer assessment and self-assessment might be encouraged if we want to maximise the effect of OPA on metacognitive thinking and reflective thinking.

The roles of OPA design elements in moderating the effect size of OPA on HOT were also analysed in this meta-analysis. These factors had no significant moderation effect, which does not align with the findings of the past meta-analyses of the effect of OPA on learning achievement. Li et al. (2020) found that rater training was a critical factor influencing the effect of peer assessment on learning achievement. Zheng et al. (2020) echoed that finding and found that moderators such as assessment task, anonymous assessment and individual peer assessment were also significant. There are four reasons that may explain the inconsistency between our findings and the finding of other meta-analyses. First, moderators such as 'training', 'assessment task' and 'grouping type' were not explored in this study due to limited variability, which made it difficult to meaningfully group the samples together (Fong et al., 2019). Second, some moderators such as 'reciprocal roles' and 'anonymity' produced a large effect size on HOT but the effect was not significant due to the small sample size. Third, multiple moderators were nested within interventions. Collinearity can produce inaccurate estimates of the real effect of individual moderators (Yan et al., 2023). Fourth, the outcome variables of other meta-analyses were academic achievement while this meta-analysis specifically focused on the outcomes of HOT.

Two moderators—reciprocal roles and anonymity—are worth special attention when designing the OPA activities to promote HOT. Although they did not cause statistically significant differences, their potential for moderating OPA effects on HOT when the sample size is larger should not be underestimated. Reciprocal roles engage students in OPA cognitively in quite different ways (Li & Grion, 2019). As feedback givers, students need to analyse peers' work by comparing it with assessment criteria, making judgements, constructing suggestions and reflecting on their own work by referring to peers' work. As feedback receivers, they should evaluate multiple points of view, reflect and assess their own work, and self-regulate their learning based on peers' comments. These reciprocal roles double students' opportunities to practise their HOT (Tsivitanidou & Constantinou, 2016). HOT may grow and be reinforced by multiple practice (Dominguez et al., 2015). Anonymity has been advocated for its potential to alleviate the undesirable interpersonal effects that can result from traditional face-to-face peer assessments (Carvalho, 2013). Anonymity enables students to constructively criticise each other's work which activate students' HOT (Lin, 2018; Raes et al., 2015).

It was interesting to find that students at school level tended to develop their HOT more than those at tertiary level via OPA, although not being statistically significant because of small sample size. This finding contradicts what Li et al. (2020) found in their meta-analysis of OPA on learning achievement. Although college students might be more cognitively competent than school students in rating each other's work (Double et al., 2020), the accuracy of peer evaluation might be not so crucial for HOT development. The younger learners seem to exceed older learners in HOT development perhaps because they are less restricted in thinking and more open-minded than the older counterparts. However, this assumption needs more empirical evidence to support. As for methodology variables, a trend is noticeable, although not statically significant, that when the research is more scientifically designed the effect size would be smaller. A less controlled experimental design and subjective evaluation of student work might inflate the effect sizes of OPA.

To conclude, we present a pioneering synthesis of the effects of OPA on HOT. Overall, OPA is strongly recommended as a powerful strategy given its significant positive effects on HOT. It was found to have a significantly larger effect on convergent HOT than on divergent HOT which could inspire practitioners to consider what they can do to promote divergent HOT via OPA. It should be noted that a total of 17 qualified papers were identified in this meta-analysis study. Although reasonable inferences could be made through analysing the 39 effect sizes reported in this paper, they are still rather hypothetical. With the accumulation of more empirical evidence in the future, the conclusions drawn in this study can be further testified. The relatively small number of existing studies on the impacts of OPA on HOT may be caused by the challenges in assessing and cultivating HOT. It normally takes longer time to cultivate HOT and is more complicated to operationalise and measure HOT than learning performance concerning factual knowledge and practical skills. Encouragingly, the publication trend shown by this study indicates the increasing research interest in this area. The results of this meta-analysis study would inspire researchers to conduct more well-controlled (quasi)experimental studies to fully examine the effects of OPA on HOT and even its effects on each type of HOT. In addition, some moderators such as training, grouping type and assessment tasks which are found as significant moderators of OPA on learning achievement by the other meta-analysis studies (Li et al., 2020; Zheng et al., 2020) were not analysed due to limited variability presented in the identified studies (Fong et al., 2019; Koenka et al., 2021). These omissions call for more research on the exploration of these factors in the mechanism of OPA on HOT.

ACKNOWLEDGEMENTS

The authors thank the editor and reviewers for the careful and insightful review and providing valuable feedback.

FUNDING INFORMATION

This research was supported by Research Cluster Fund (No: RG76/2020-2021R) in the Education University of Hong Kong.

CONFLICT OF INTEREST STATEMENT

No potential conflict of interest was reported by the authors.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no new data were generated or analysed during the current study.

ETHICS STATEMENT

This meta-analysis does not involve direct data collection from human participants.

ORCID

Zhi Hong Wan  <https://orcid.org/0000-0002-2362-2887>

Lan Yang  <https://orcid.org/0000-0002-3457-0330>

REFERENCES

- Acuna, E., & Rodriguez, C. (2004). *A meta-analysis study of outlier detection methods in classification*. Department of Mathematics, University of Puerto Rico at Mayaguez. https://www.researchgate.net/publication/228728761_A_meta_analysis_study_of_outlier_detection_methods_in_classification
- Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science*, 29(1), 1–32. <https://doi.org/10.1023/A:1026471702353>
- Baker, M., Rudd, R., & Pomeroy, C. (2001). Relationships between critical and creative thinking. *Journal of Southern Agricultural Education Research*, 51(1), 173–188.
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41(2), 257–278. <https://doi.org/10.1111/j.2044-8317.1988.tb00901.x>
- Beyer, B. K. (1987). *Practical strategies for the teaching of thinking*. Allyn and Bacon.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Brookhart, S. M. (2018). Summative and formative feedback. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 52–78). Cambridge University Press. <https://doi.org/10.1017/9781316832134.005>
- Bruner, J. (1973). *The relevance of education*. W.W. Norton & Company.
- Bruning, R. H., Schraw, G. J., Norby, M. N., & Ronning, R. R. (2004). *Cognitive psychology and instruction* (5th ed.). Pearson.
- Carvalho, A. (2013). Students' perceptions of fairness in peer assessment: Evidence from a problem-based learning course. *Teaching in Higher Education*, 18(5), 491–505. <https://doi.org/10.1080/13562517.2012.753051>
- Chen, T. (2016). Technology-supported peer feedback in ESL/EFL writing classes: A research synthesis. *Computer Assisted Language Learning*, 29(2), 365–397. <https://doi.org/10.1080/09588221.2014.960942>
- Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education*, 25, 78–84. <https://doi.org/10.1016/j.iheduc.2015.02.001>
- Cheng, M. H. M., & Wan, Z. H. (2017). Exploring the effects of classroom learning environment on critical thinking skills and disposition: A study of Hong Kong 12th graders in Liberal studies. *Thinking Skills and Creativity*, 24, 152–163. <https://doi.org/10.1016/j.tsc.2017.03.001>
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338. <https://doi.org/10.1016/j.learninstruc.2009.08.006>
- Ciftci, H., & Kocoglu, Z. (2012). Effects of peer e-feedback on Turkish EFL students' writing performance. *Journal of Educational Computing Research*, 46(1), 61–84. <https://doi.org/10.2190/EC.46.1.c>
- Cockett, A., & Jackson, C. (2018). The use of assessment rubrics to enhance feedback in higher education: An integrative literature review. *Nurse Education Today*, 69, 8–13. <https://doi.org/10.1016/j.nedt.2018.06.022>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Cowan, J. (2010). Developing the ability for making evaluative judgements. *Teaching in Higher Education*, 15(3), 323–334. <https://doi.org/10.1080/13562510903560036>
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. Harper Collins.
- Dominguez, C., Nascimento, M. M., Payan-Carreira, R., Cruz, G., Silva, H., Lopes, J., Morais, M. D. F. A., & Morais, E. (2015). Adding value to the learning process by online peer review activities: Towards the elaboration of a methodology to promote critical thinking in future engineers. *European Journal of Engineering Education*, 40(5), 573–591. <https://doi.org/10.1080/03043797.2014.987649>
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32(2), 481–509. <https://doi.org/10.1007/s10648-019-09510-3>
- Filius, R. M., de Kleijn, R. A., Uijl, S. G., Prins, F. J., van Rijen, H. V., & Grobbee, D. E. (2018). Strengthening dialogic peer feedback aiming for deep learning in SPOCs. *Computers & Education*, 125, 86–100. <https://doi.org/10.1016/j.compedu.2018.06.004>
- Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. *Educational Psychology Review*, 31(1), 121–162. <https://doi.org/10.1007/s10648-018-9446-6>

- Gezie, A., Khaja, K., Chang, V. N., Adamek, M. E., & Johnsen, M. B. (2012). Rubrics as a tool for learning and assessment: What do baccalaureate students think? *Journal of Teaching in Social Work*, 32(4), 421–437. <https://doi.org/10.1080/08841233.2012.705240>
- Gikandi, J. W., & Morrow, D. (2016). Designing and implementing peer formative feedback within online learning environments. *Technology, Pedagogy and Education*, 25(2), 153–170. <https://doi.org/10.1080/1475939X.2015.1058853>
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, 115(4), 523–547. <https://doi.org/10.1086/681947>
- Guiller, J., Durnell, A., & Ross, A. (2008). Peer interaction and critical thinking: Face-to-face or online discussion? *Learning and Instruction*, 18(2), 187–200. <https://doi.org/10.1016/j.learninstruc.2007.03.001>
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781003107347>
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.3102/10769986006002107>
- Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486–504. https://faculty.ucmerced.edu/jvevea/classes/290_21/readings/week%2011/Hedges%20and%20Veevea%201998.pdf
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.
- Huang, Y. M., Silitonga, L. M., & Wu, T. T. (2022). Applying a business simulation game in a flipped classroom to enhance engagement, learning achievement, and higher-order thinking skills. *Computers & Education*, 183, e104494. <https://doi.org/10.1016/j.compedu.2022.104494>
- Ketonen, L., Nieminen, P., & Häikiöniemi, M. (2020). The development of secondary students' feedback literacy: Peer assessment as an intervention. *The Journal of Educational Research*, 113(6), 407–417. <https://doi.org/10.1080/00220671.2020.1835794>
- Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2021). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback. *Educational Psychology*, 41(7), 922–947. <https://doi.org/10.1080/01443410.2019.1659939>
- Leader, D. C., & Clinton, M. S. (2018). Student perceptions of the effectiveness of rubrics. *Journal of Business and Educational Leadership*, 8(1), 86–103. http://asbbs.org/files/2019/JBEL_8.1_Fall_2018.pdf#page=86
- Lewis, A., & Smith, D. (1993). Defining higher-order thinking. *Theory into Practice*, 32(3), 131–137. <https://doi.org/10.1080/00405849309543588>
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywonwi, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(2), 193–211. <https://doi.org/10.1080/02602938.2019.1620679>
- Li, L., & Grion, V. (2019). The power of giving feedback and receiving feedback in peer assessment. *All Ireland Journal of Higher Education*, 11(2), 1–17. <https://ojs.aishe.org/index.php/aishe-j/article/view/413/671>
- Li, L., Liu, X., & Zhou, Y. (2012). Give and take: A re-analysis of assessor and assessee's roles in technology-facilitated peer assessment. *British Journal of Educational Technology*, 43(3), 376–384. <https://doi.org/10.1111/j.1467-8535.2011.01180.x>
- Lin, G. Y. (2018). Anonymous versus identified peer assessment via a Facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers & Education*, 116, 81–92. <https://doi.org/10.1016/j.compedu.2017.08.010>
- Lin, H. C., Hwang, G. J., Chang, S. C., & Hsu, Y. D. (2021). Facilitating critical thinking in decision making-based professional training: An online interactive peer-review approach in a flipped learning context. *Computers & Education*, 173, e104266. <https://doi.org/10.1016/j.compedu.2021.104266>
- Liu, J., Ma, Y., Sun, X., Zhu, Z., & Xu, Y. (2021). A systematic review of higher-order thinking by visualizing its structure through HistCite and CiteSpace software. *The Asia-Pacific Education Researcher*, 11, 635–645. <https://doi.org/10.1007/s40299-021-00614-5>
- Lu, K., Pang, F., & Shadiev, R. (2022). How to deepen college students' approach to using technologies in T-O-IBL? Examining the mediating influence of deep approaches to using technologies between learning factors and higher order thinking skills. *Journal of Computer Assisted Learning*, 12, 182–193. <https://doi.org/10.1111/jcal.12738>
- 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(4), 353–369. <https://doi.org/10.1007/s11165-006-9029-2>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *Annals of Internal Medicine*, 151(4), 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

- Newmann, F. M. (1991). Promoting higher order thinking in social studies: Overview of a study of 16 high school departments. *Theory & Research in Social Education*, 19(4), 324–340. <https://doi.org/10.1080/00933104.1991.10505645>
- Noroozi, O., & Mulder, M. (2017). Design and evaluation of a digital module with guided peer feedback for student learning biotechnology and molecular life sciences, attitudinal change, and satisfaction. *Biochemistry and Molecular Biology Education*, 45(1), 31–39. <https://doi.org/10.1002/bmb.20981>
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23–38. <https://doi.org/10.1080/02602930050025006>
- Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education*, 44(8), 1253–1278. <https://doi.org/10.1080/02602938.2019.1600186>
- Panadero, E., Jonsson, A., & Alqassab, M. (2018). Providing formative peer feedback: What do we know? In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 409–431). Cambridge University Press. <https://doi.org/10.1017/9781316832134>
- Patchan, M. M., Schunn, C. D., & Clark, R. J. (2018). Accountability in peer assessment: Examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education*, 43(12), 2263–2278. <https://doi.org/10.1080/03075079.2017.1320374>
- Prayoonsri, B., Tatsirin, S., Suntorapot, D., & Jariya, C. (2015). Factors affecting higher order thinking skills of students: A meta-analytic structural equation modeling study. *Educational Research and Reviews*, 10(19), 2639–2652. <https://doi.org/10.1080/03054980701476386>
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), 1–20. <https://doi.org/10.1080/03054980701476386>
- Raes, A., Vanderhoven, E., & Schellens, T. (2015). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education*, 40(1), 178–193. <https://doi.org/10.1080/03075079.2013.823930>
- Resnick, L. B. (1987). *Education and learning to think*. National Academy Press. <http://elibrary.pcu.edu.ph:9000/digi/NA02/1987/1032.pdf>
- Richmond, A. S., & Hagan, L. K. (2011). Promoting higher level thinking in psychology: Is active learning the answer? *Teaching of Psychology*, 38(2), 102–105. <https://doi.org/10.1177/0098628311401581>
- Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice* (2nd ed.). Elsevier Academic Press.
- Schwarzer, G. (2007). Meta: An R package for meta-analysis. *R News*, 7(3), 40–45. https://cran.rstudio.org/doc/Rnews/Rnews_2007-3.pdf#page=40
- Shang, H. F. (2017). An exploration of asynchronous and synchronous feedback modes in EFL writing. *Journal of Computing in Higher Education*, 29(3), 496–513. <https://doi.org/10.1007/s12528-017-9154-0>
- Shin, Y., Kim, D., & Song, D. (2020). Types and timing of scaffolding to promote meaningful peer interaction and increase learning performance in computer-supported collaborative learning environments. *Journal of Educational Computing Research*, 58(3), 640–661. <https://doi.org/10.1177/0735633119877134>
- Sosu, E. M. (2013). The development and psychometric validation of a critical thinking disposition scale. *Thinking Skills and Creativity*, 9, 107–119. <https://doi.org/10.1016/j.tsc.2012.09.002>
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>
- Tanujaya, B., Mumu, J., & Margono, G. (2017). The relationship between higher order thinking skills and academic performance of student in mathematics instruction. *International Education Studies*, 10(11), 78–85. <https://doi.org/10.5539/ies.v10n11p78>
- Tenório, T., Bittencourt, I. I., Isotani, S., & Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A systematic review of the literature. *Computers in Human Behavior*, 64, 94–107. <https://doi.org/10.1016/j.chb.2016.06.020>
- Tishman, S., Jay, E., & Perkins, D. N. (1993). Teaching thinking dispositions: From transmission to enculturation. *Theory into Practice*, 32(3), 147–153. <https://doi.org/10.1080/00405849309543590>
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. <https://doi.org/10.3102/00346543068003249>
- Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 55–87). Springer. https://doi.org/10.1007/0-306-48125-1_4
- Tsvitanidou, O. E., & Constantinou, C. P. (2016). A study of students' heuristics and strategy patterns in web-based reciprocal peer assessment for science learning. *The Internet and Higher Education*, 29, 12–22. <https://doi.org/10.1016/j.iheduc.2015.11.002>

- Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education*, 31(3), 341–356. <https://doi.org/10.1080/03075070600680836>
- van Gennip, N. A., Segers, M. S., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, 4(1), 41–54. <https://doi.org/10.1016/j.edurev.2008.11.002>
- van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24–34. <https://doi.org/10.1016/j.edurev.2016.10.003>
- van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270–279. <https://doi.org/10.1016/j.learninstruc.2009.08.004>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, 100(4), 930. <https://doi.org/10.1037/a0012842>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Yan, Z., Wang, X., Boud, D., & Lao, H. (2023). The effect of self-assessment on academic performance and the role of explicitness: A meta-analysis. *Assessment & Evaluation in Higher Education*, 48(1), 1–15. <https://doi.org/10.1080/02602938.2021.2012644>
- Yen, T. S., & Halilli, 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://tojedel.net/journals/tojedel/articles/v03i02/v03i02-04.pdf>
- Zhan, Y. (2021). What matters in design? Cultivating undergraduates' critical thinking through online peer assessment in a Confucian heritage context. *Assessment & Evaluation in Higher Education*, 46(4), 615–630. <https://doi.org/10.1080/02602938.2020.1804826>
- Zhan, Y., Wan, Z. H., & Sun, D. (2022). Online formative peer feedback in Chinese contexts at the tertiary level: A critical review on its design, impacts and influencing factors. *Computers & Education*, 176, e104341. <https://doi.org/10.1016/j.compedu.2021.104341>
- Zhang, Y., Pi, Z., Chen, L., Zhang, X., & Yang, J. (2021). Online peer assessment improves learners' creativity: Not only learners' roles as an assessor or assessee, but also their behavioral sequence matter. *Thinking Skills and Creativity*, 42, e100950. <https://doi.org/10.1016/j.tsc.2021.100950>
- Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(3), 372–386. <https://doi.org/10.1080/02602938.2019.1644603>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology*, 54, 817–835. <https://doi.org/10.1111/bjet.13310>