Thinking Skills and Creativity 49 (2023) 101362

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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

It is not what you think it is how you think: A critical thinking intervention enhances argumentation, analytic thinking and metacognitive sensitivity

Boban Simonovic[a,*], Katia Vione[a], Edward Stupple[a], Alice Doherty[b]

[a] University of Derby, Kedleston Road, Derby DE22 1GB UK
[b] University of Nottingham, UK

ARTICLE INFO

Keywords:
Analytic writing
Argument evaluation
Cognitive reflection
Critical thinking
Meta-reasoning

ABSTRACT

Critical thinking (CT) is an essential element of Higher Education and is central to graduate achievement and employability. This study examined the effect of an online CT intervention (developed using dual process and meta-reasoning theoretical frameworks) on Higher Education students' attitudes about CT and their CT skills. One hundred and forty-eight participants were allocated either to an intervention or control group, assessed at pre and post-test for perceptions and attitudes towards CT, cognitive reflection, argument evaluation, and analytical writing style (via an essay writing task). The analyses results revealed a significant increase from pre- to post-intervention on all CT measures except for analytic writing style and valuing critical thinking, with the intervention group performing significantly better at post-test than the control group. It was concluded that CT can be improved with brief online interventions based on ''how to think'' rather than ''what to think'' and that attitudes and beliefs play an important part in the development of students' CT skills.

1. Introduction

Critical thinking (CT) is vital for success in higher education as it plays a central role in problem-solving, thinking and decision-making (Dwyer et al., 2014; Ennis, 2003; Facione, 1990; Halpern, 2014; Puig et al., 2019). CT skills facilitate students' learning which is essential to their academic performance (Davies, 2006). Research also associates good CT skills with career advancement, higher rates of employability and generation of innovative solutions (Facione & Facione, 2001; Osborne et al., 2013).

Developing CT skills is increasingly essential, not only in academic settings but also for adapting to a changing world (Halpern, 2003). Future graduates will need to solve problems which do not yet exist, with CT skills essential for equipping individuals to be flexible, adaptive and change resilient (Cloete, 2018). Good critical thinkers make better decisions (Gambrill, 2006), are less susceptible to cognitive biases and fallacies (Stanovich & West, 2008), and are more likely to use an analytic writing style associated with logical and evaluative thinking (Pennebaker et al., 2015). However, many students struggle to understand CT and are unsure how to develop their CT skills which consequently leads to a lack of confidence in its application (Duro et al., 2013) or a disconnect between confidence and performance (Kruger & Dunning, 1999).

There have been challenges in defining CT (Tsui, 2002); the American Philosophical Association expert consensus defined CT as:

* Corresponding author.
E-mail address: b.simonovic1@derby.ac.uk (B. Simonovic).

https://doi.org/10.1016/j.tsc.2023.101362
Received 24 November 2022; Received in revised form 15 May 2023; Accepted 22 June 2023
Available online 26 June 2023

"purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (Facione, 1990, p. 2). This definition indicates the complex nature of CT, and the challenge of developing CT interventions (Dwyer et al., 2014; Ku, 2009; Tsui, 2002). Dwyer et al. (2014) proposed an integrative framework that define CT as a metacognitive process that consists of analysis, inference, and the drawing of conclusions whereby the self-regulation processes and disposition towards thinking are fundamental in decision-making, problem solving and reasoning (Dwyer et al., 2014; Halpern, 2003, 2014).

Dwyer's framework can be elaborated using meta-reasoning theory (Ackerman & Thompson, 2017) which postulates that metacognitive monitoring (as measured by feelings of rightness or confidence in responses given) is important for thinking and reasoning performance through its influence on metacognitive control (which includes the decision to provide a response, switch strategy or give up). Ackerman and Thompson (2017) argued that Meta-reasoning involves the ability to reflect on and analyze one's own reasoning processes. This includes being able to identify assumptions, evaluate evidence, and consider alternative perspectives. An important aspect of the theory is the potential for directing participants to monitor their cognitions to facilitate metacognitive awareness of reasoning processes to help identify potential thinking errors, and encourage the deployment of alternative strategies for CT (Thompson et al., 2013). Thus, by teaching individuals to engage in meta-reasoning, we can help them become more self-aware and reflective thinkers, which can improve their overall CT skills and ameliorate their propensity for bias.

Within this framework metacognitive monitoring processes are important to enhancing CT. This involves being able to assess the quality of one's own reasoning, identify potential errors or biases and develop a confidence in providing adequate solutions to a problem. Before engaging in problem solving, individuals engage in a subjective assessments of the probability of success or a failure (initial judgment of solvability) followed by assessing feeling of rightness (i.e. whether an initial response is correct) and feeling of error (trying to identify whether there is a mistake in response) which determine the final confidence in their response (Ackerman, 2014; Ackerman & Thompson, 2017; Gangemi et al., 2015; Thompson & Johnson, 2014). Hence, encouraging metacognitive monitoring can help individuals develop effective strategies for evaluating evidence, assessing arguments, and making sound judgments. The meta-reasoning framework originates from the meta-memory literature (e.g., Koriat & Goldsmith, 1996) with parallels between judgements of learning and feeling of rightness correlating with response accuracy (e.g., Nelson & Narens, 1990). Note that in the present paper we do not seek to make fundamental claims about metacognition or unpick the nuanced elements of metacognitive awareness, knowledge or skill. Instead, we use post answer confidence to examine whether an intervention based upon dual process theory and the meta-reasoning framework can improve metacognitive sensitivity and task performance.

The meta-reasoning framework supports the theoretical positions of Stanovich (2016) and Bonnefon (2018), with clear overlap between the framework and the characteristics of the analytic system postulated by dual-process theories. Dual-process theories of cognitive processing differentiate between two processes: "Type 1″ which is fast, automatic, and intuitive, but is prone to systematic errors and miserly processing of information, and "Type 2″ which entails slow, deliberate, self-regulatory analytic processing that is taxing for working memory (e.g., Toplak et al., 2014). Cognitive and metacognitive components of the Type 2 processes may be pivotal in the development and monitoring of appropriate "mindware" (rules and explicit procedures) that enable the learning and application of CT skills (Bonnefon, 2018).

There is a broad range of practise across disciplines, and depending on the subject matter, generic or content-specific approaches are used for teaching CT (Abrami et al., 2008). For example, there is evidence that CT requires mastery of the content-specific knowledge in the subject area (Bailin & Siegel, 2003) but also evidence that CT relates to more general, non-specific knowledge (Barrie, 2006). Arguably, these varying approaches make it difficult for teachers and researchers to understand and agree on the focal points and key components of CT to teach and research. Despite potential difficulties related to training and teaching of CT skills, several meta-analyses indicate that the most effective interventions are instructional interventions (Abrami et al., 2008, 2015; Puig et al., 2019; Tiruneh et al., 2014). Specifically, a mixed approach is most effective and entails a combination of the general approach (teaching skills and dispositions without subject knowledge) with either infusion (thinking critically on the subject matter) or immersion approaches (subject matter is thought-provoking) (Abrami et al., 2008; 2016; Ennis, 2016; Puig et al., 2019). This is the CT intervention approach employed in the present study.

Previous research reveals that CT can be improved through interventions ranging from 4-weeks (e.g., Cloete, 2018; Ghanizadeh, 2017; Simonovic et al., 2022) to 8 weeks (e.g., Dwyer & Walsh, 2020, 2007). The measures used in such interventions were primarily focused on reasoning and problem-solving (e.g., the Watson-Glaser Critical Thinking Appraisal, Watson & Glaser, 1964; Halpern Critical Thinking Assessment, Halpern, 2010) or thinking dispositions (e.g., Need for Cognition Scale, Cacioppo et al. 1984; reflective thinking questionnaire Kember et al., 2000) rather than CT skills, thus relying on formal reasoning tasks that are somewhat limited in scope The analysis and evaluation of skills and thinking dispositions are most used and some authors focus on the skills such as interpretation and analysis (e.g., Vertecchi et al., 2017), thus neglecting other metacognitive processes such as self-regulation, control, monitoring and reflective judgement (Dwyer et al., 2012, 2014; Dwyer and Walsh, 2020).

Indeed, Type 2 thinking and meta-reasoning processes, such as self-regulation, monitoring and reflective thinking are important facets of CT (Ackerman & Thompson, 2017; Bonnefon, 2018; Halpern, 2014; Magno, 2010; Stanovich, 2016). This supports an integrative framework of CT that entails metacognitive, reflective, self-regulatory processes (Ackerman & Thompson, 2017; Dwyer et al. 2014), whereby thinking disposition may offer an important perspective on the development of CT skills. Indeed, ample evidence indicates that thinking dispositions, beliefs and skills are necessary for the development of CT (Ferguson & Bråten, 2013; Huber & Kuncel, 2016; Stupple et al., 2017a). However, thinking dispositions need to be relevant to the topic under consideration (in this case Higher Education), thus it is important to capture students' beliefs and attitudes about CT (Stupple et al., 2017a). Simonovic et al. (2022) found that student attitudes and beliefs have the effect to motivate students to think about their CT skills and abilities. Furthermore, a CT intervention in their study significantly improved cognitive reflection, student confidence and beliefs about CT and their grade point average (GPA). This indicates that CT confidence and beliefs can activate or deactivate CT skills. It is therefore beneficial to exploit a wide range of measures to examine the multifaceted nature of CT, including measures of beliefs and attitudes as well as CT behaviours.

Given that CT can be taught, and developing CT skills is largely dependent on the method in which CT instructions are delivered (Abrami et al., 2008, 2015; Halpern 2014; Dwyer et al. 2012), it is vital to understand the effectiveness of CT interventions in facilitating improvements across the various facets of CT. Therefore, a CT intervention developed previously (see Simonovic et al., 2022) was used to encompass these aspects of CT (Ennis, 2016). This intervention was based upon dual-process and meta-reasoning theories and included elements that have been shown to be successful in teaching CT in meta-analytical studies (Abrami et al., 2008, 2015; Puig et al., 2019; Tiruneh et al., 2014). This novel approach empowers students to apply CT skills in a wide range of contexts by instilling CT skills through real life-examples that stimulates shift between intuitive and deliberative processing via cognitive and metacognitive processes whereby thought-provoking ideas provide immersion in subject matter (e.g., argument evaluation).

The intervention addressed students' understanding of CT and aimed to enhance this for application in the real world by showing students they need to know more than just what, but also why and how when engaging in CT. Specifically, the focus was on the enhancement of CT through evaluating assumptions (i.e., breaking down complex information into smaller parts and assessing the strengths and weaknesses of an argument), inferences (i.e., drawing conclusions based on evidence) and observation (e.g., Halpern, 2003). The intervention further focused on metacognitive monitoring and control to increase confidence during the analytic processing of tasks (e.g., Ackerman & Thompson, 2017; Stanovich, 2018). We used real-life examples to "sharpen" students' metacognitive skills by demonstrating how to think about common errors in reasoning and how to avoid biases (e.g., Kahneman & Tversky, 1984; Tversky & Kahneman, 1983). Finally, the intervention focused on significant aspects of dual-process theory: biases, fallacies, and heuristics that can highlight the need for deliberative thinking (Evans & Stanovich, 2013) with examples based upon fake news and beliefs based on social media including argument evaluation and evidence evaluation tasks. Calvillo et al. (2022) demonstrated that brief interventions can enhance cognitive reflection by encouraging the consideration of alternative responses, they did not however find evidence that these interventions generalised to more other thinking tasks. Nevertheless, the present intervention further targets meta-reasoning processes and participant beliefs and attitudes. The goal of the intervention was to facilitate students to become critical thinkers using a "how to think" teaching approach which allows students to learn from real-life examples and to apply principles of dual-processing and meta-reasoning theories to reflect on their thought processes and potential for error and bias Thus, the intervention was designed to encourage, stimulate, and facilitate the acquisition of CT skills by providing the necessary mindware such as awareness of common fallacies and biases (Abrami et al., 2015; Bonnefon, 2018; Stanovich, 2009), emphasis on the generation of alternative answers/solutions, encouragement to slow down to avoid biases as well as reflecting on their confidence in their task performance.

The present study expands on Simonovic et al. (2022), testing the effect of an intervention on several measures of CT ability, dispositions and behaviours that map onto Type 2 processes. We used text analysis to explore changes in students' linguistic style. Specifically, we used the Analytic Thinking measure from Linguistic Inquiry and Word Count (LIWC): a well-established and validated text analysis program with an algorithm for measuring individual differences in analytic cognitive style. Originally tested as a metric for predicting academic success (Pennebaker et al., 2014), the Analytic Thinking measure (or Categorical Dynamic Index as it has been known) is calculated based on word frequency counts in linguistic samples. To the authors' knowledge, the link between Analytic Thinking scores and CT abilities and behaviours has not been tested, however, the Analytic Thinking measure has been shown to correlate positively with Scholastic Aptitude Tests (Pennebaker et al., 2014) and has been shown to be a predictor of academic success in higher education (e.g., Abe, 2020; Pennebaker et al., 2014; Pennebaker & King, 1999). The current research will therefore employ this measure to explore whether the CT intervention leads to incidental changes in academic writing style.

We used the Critical Thinking Toolkit (CriTT; Stupple. et al., 2017a) to measure potential changes in CT attitudes and beliefs. The CriTT entails "confidence in critical thinking", "valuing critical thinking" and "misconception of critical thinking", factors which correlate with theoretical and applied aspects of CT. Deliberative and intuitive thinking were examined with the Cognitive Reflection Test (CRT; Toplak et al., 2014) and the Argument Evaluation Test (AET; Stanovich & West, 1997). The seven-item CRT tests the ability to inhibit impulsive judgments to engage reflective thinking processes and correlates with measures of cognitive ability (e.g., 0.50 correlations with Wechsler Abbreviated Scale of Intelligence) and thinking disposition measures (e.g., 0.41 with Active Open-Minded Thinking) (Toplak et al., 2014). Thus, the CRT has a strong predictive capability because it captures two important aspects of reasoning: cognitive ability and thinking disposition. Metacognitive monitoring was assessed with the CRT by correlating final confidence with success across items. Two important aspects of argumentation, fundamental to CT - the influence of beliefs on judgements of argument strength and the extent to which judgements of argument strength align with expert judgements were measured with the AET. The pitfalls of belief bias were highlighted during the intervention and as such we predicted participant performance to improve on the AET.

The proposal that writing style and CT are related has been evidenced in past research (see e.g., Carroll, 2007) but the Analytic Thinking measure in LIWC has yet to be linked to the development of CT skills. The Analytic Thinking measure positively weights number of articles (e.g. a, the, some) which have been associated with abstract thinking, and number of prepositions (e.g. in, with, because of, before) which have been associated with cognitive complexity, and it negatively weights the following word classes that have been linked with a more dynamic, narrative language style: auxiliary verbs (can, must); adverbs (often, happily); conjunctions (and, but); personal/impersonal pronouns (he, she); and negations (no, not) (Pennebaker et al. 2014). As highlighted, analytic thinking has been consistently shown to correlate with academic success (e.g., Abe, 2020; Pennebaker et al., 2014; Pennebaker & King, 1999). A high score indicates a categorical writing and thinking style that is more formal and logical and a low score indicates a more informal, intuitive writing and thinking style that is focused on the here-and-now (Boyd & Pennebaker, 2015). Increases in Analytic Thinking scores from pre and post-test would indicate changes in the linguistic markers associated with abstract thinking and cognitive complexity, which tends to be rewarded in higher education (Abe, 2020). It was hypothesised that participants' cognitive reflection (CRT), confidence in CRT, metacognitive sensitivity, valuing CT (CriTT), argument driven response (AET) and LIWC analytic thinking would increase from pre-post testing. We further hypothesised that participants' misconceptions of CT (CriTT) and belief driven responses will decrease from pre-post testing. Thus, we expected an increase in all utilised measures after the intervention for the intervention group except for the misconception scale and belief-driven response, where a decrease was expected.

2. Methods

2.1. Participants

A total of 148 participants (age, $M = 27.23$ years, $SD = 6.08$; 56.08% female) took part in the study. Participant inclusion criteria were native English speakers, aged at least 18 years, and currently studying an undergraduate or postgraduate course in the UK. Participants gave written informed consent in line with local university and British Psychological Society ethical guidelines. Participants reported their age, sex, and their study level. Participants were randomly assigned to groups and had similar demographics and study levels (Experimental group, age, $M = 26.52$ years, $SD = 6.27$; 63% female; study level, *Level 5* = 33, *Level 6* = 26, *Level 7* = 21; Control group, age, $M = 28.05$ years, $SD = 5.79$; 49% female; study level, *Level 5* = 23, *Level 6* = 23, *Level 7* = 22). Participants completed three parts of the study (pre-and-post testing and intervention) if they were in the intervention group ($N = 80$) or two parts (pre-and-post testing) if they were in the control group ($N = 68$). Participants received compensation of £2.50 per part (a total of £7.50 for the intervention group and £5 for the control group).

2.2. Materials

2.2.1. The critical thinking toolkit

The 27-item CriTT (Stupple et al., 2017a) measures student beliefs, attitudes, and perceptions of CT across three factors. *Confidence* (e.g., "I can identify the structure of arguments without being distracted by their content"); *Valuing* (e.g., "Critical thinking is essential in higher education"); and *Misconception* ("Critical thinking is when you describe what is wrong with something"). Each item is rated on a Likert scale from "1 - Strongly disagree" to "10 - Strongly Agree" (Cronbach's $\alpha = 0.740$). The subscales were analysed separately to explore the potential that they could make separate contributions to the understanding of variance in CT. Reliability measures in the present study: Confidence CriTT, Cronbach's $\alpha = 0.760$; Valuing CriTT, Cronbach's $\alpha = 0.771$; Misconception CriTT, Cronbach's $\alpha = 0.716$. Higher scores on Confidence CriTT indicate higher confidence in CT; higher scores on Valuing CriTT indicate higher recognition of the importance of CT, and higher scores on Misconception CriTT indicate avoidance of CT or higher misconception of CT.

2.2.2. The extended cognitive reflection test

Toplak et al.'s (2014) Extended CRT comprises 7 mathematical reasoning items to measure the ability to resist intuitive responses to engaging analytic thinking (e.g., "A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?"). The intuitive response is to respond that the ball costs $0.10 (totalling $1.20), but the correct answer is $0.05. Each item is scored as either correct or incorrect. Higher scores indicate greater levels of cognitive reflection (Cronbach's $\alpha = 0.720$).

2.2.3. CRT 2

The four items CRT-2 (Thomson & Oppenheimer, 2016) measures cognitive reflection but reduces the role of numeracy compared to the original CRT. An example item is: "If you're running a race and you pass the person in second place, what place are you in?" (Intuitive answer: First; correct answer: second). Items are scored as correct or incorrect and a total is derived. Higher scores indicate increased cognitive reflection (Cronbach's $\alpha = 0.720$). We mitigated the effects of mathematical ability by including two versions of the CRT, with counterbalanced items presented in the pre and post-intervention conditions.

2.2.4. Confidence in CRT

We measured confidence after each CRT question on a 10-point Likert scale to examine metacognitive processes in responses to CRT questions (Cronbach's $\alpha = 0.785$). Additionally, we correlated the mean CRT confidence with a percentage of correct CRT responses to establish a metacognitive sensitivity measure (see e.g., Ackerman et al., 2016).

2.2.5. Argument evaluation test

Stanovich and West's (1997) Argument Evaluation Test (AET, modified to a 19-item version suitable for a UK population) was used to measure participants' ability to objectively evaluate argument quality. In the first phase of this task, participants rated the extent of their agreement with statements about controversial topics (e.g., "The national debt should be reduced by cutting politicians' salaries"). This phase measures belief-driven responding (Cronbach's $\alpha = 0.747$).

In the argument evaluation phase of the task, three elements were presented, for example, *Claim: "Politicians' salaries are very high and cutting them would make a significant step towards paying off the huge national debt." 2. Counterargument: "The national debt is so large that totally eliminating all politicians' salaries would still hardly make a dent in it." 3. Rebuttal to counterargument: "The politicians, whose actions are to a considerable extent responsible for the huge national debt, earn salaries several times higher than the national average."*. Participants then rated the quality of the argument presented in the rebuttal on a four-point Likert scale. This phase measures argument-driven responding (Cronbach's $\alpha = 0.739$).

Regression analysis for each participant examined the extent to which belief and argument strength predicted their argument evaluations. This was achieved using beta values derived from the relationship between their beliefs (agreement ratings) and argument ratings, and the extent to which participant argument ratings were predicted by the expert panel ratings from the original Stanovich & West (1997) study. The beta values indicated (i) the extent to which participants' judgments about argument strength aligned with their own pre-existing beliefs (indicative of a preference for intuitive thinking); and (ii) the extent to which they suppress the influence of their beliefs to align with expert argument ratings (indicative of a preference for analytic thinking).

2.2.6. *Analytical thinking scale (ATS; Pennebaker, 2015)*

Participants were tasked to write a 500-word essay to allow evaluation of their analytical thinking, which was analysed using the procedure adopted by Pennebaker et al. (2015) and the software LIWC. The Analytical Thinking variable is derived from a factor analysis of measures of word categories used in text (personal pronouns, impersonal pronouns, auxiliary verbs, articles, prepositions, conjunctions, negations, and common adverbs). High score on Analytical Thinking variable indicates high reasoning skills, logical and hierarchical thinking patterns. In Phase 1, participants were tasked to write a short essay answering the question "Describe an accomplishment, event, or realization that sparked a period of personal growth and a new understanding of yourself or others. You may wish to consider why it was so important? What did you learn from it?". In Phase 2, the essay instructions were "Describe a topic, idea, or concept you find so engaging that it makes you lose all track of time. Why does it captivate you? What or who do you turn to when you want to learn more?". Based on their response, analytical writing style variables were created.

2.2.7. *Intervention*

The 45-min (3 parts of 15 min) intervention was designed and created by the research team using a screen and voice capture software to create short video presentation of the instructional element of the intervention. The videos were then embedded in Qualtrics survey software to present a series of short tasks, monitor engagement, and record participant responses. The aim of the intervention was to enhance CT by showing participants how to engage with observations, inferences, tests of assumptions; consideration of fallacies, biases, and heuristics in reasoning and CT and techniques to evaluate arguments and evidence evaluation (e.g., Halpern, 2003; Kahneman & Tversky, 1984; Tversky & Kahneman, 1983).

The intervention combined text and audio material. To support task engagement, the participants could not proceed to the next section until a task calibrated time period had expired. All parts of the intervention followed the same pattern. Participants were first asked to complete an activity (based on a real-life scenario) whereby they engaged in decision making processes to solve a problem. Following, the activity was explained, the theories that account for decision-making steps and reasoning were explained and participants needed to reflect on their thought processes. This required participants to recognize the existence of a problem and an acceptance of the need for evidence in support of valid inferences. Participants then engaged with metacognitive training including task monitoring, reflection and deriving logical conclusion. All examples and scenarios were explained and linked to dual-process and meta-reasoning theories (see supplementary Appendix). For example, Part 1, Activity 1 was devised to encourage participants to pause for thought to both engage analytic thinking and consider alternative explanations (based upon dual-process theory). In contrast, Part 2, Activity 2 was devised to prompt participants to 'think about their thinking' – the introduction of metacognitive concepts of knowledge of cognition and regulation of cognition were included to encourage reflection and for participants to explicitly consider their monitoring and control of these processes (based upon meta-reasoning theory).

2.3. *Procedure*

Participants were recruited via Prolific (the online platform). For the intervention group, the study consisted of three parts. In Part 1 (pre-test), participants completed the baseline online survey, including demographic information and the measures described above in randomized order and were tasked to write a the 500-word essay. Part 2 (intervention) consisted of an adaptation of the intervention utilised in Simonovic et al. (2022), described previously. Participants completed this task in one session, which was sent via email 24 h after they completed Part 1. Finally, Part 3 (post-test) was completed 7 days after participants completed Part 2. This follow-up survey asked participants to complete the same questionnaires in a randomized order as in Part 1 (pre-test) and were tasked to write another 500 words essay. For the control group, the procedure was repeated except participants did not engage with the intervention (Part 2). After completion of all parts of the study, debriefing material was provided to all participants.

2.4. *Analytic strategy*

An a priori power analysis was conducted estimating a small to medium effect size (0.15), power 0.80, and alpha level set at $p < .05$, which resulted in a total sample of 148 participants. Data were screened using graphical and statistical methods to establish suitable methods for analysis. A series of one-way of covariance (ANCOVA) on post scores, with baseline scores as a covariate, were conducted on all measures (CRT, CRT Confidence, Metacognitive sensitivity; CriTT subscales-Confidence, Valuing and Misconception; AET-Belief Driven Responses; BDR; AET-Argument-Driven Responses; ADR and ATS-Analytic writing) with condition (intervention, control) as the between-subjects factor. Finally, we conducted a series of post-hoc tests to examine the differences between the pre- and post-test scores in the experimental group.

3. Results

3.1. Comparison across conditions and time

Descriptive statistics for all variables across Time and Condition are shown in Table 1.

A multiple, one-way analysis of covariance (ANCOVA) was run to examine whether there were differences in all scores between the intervention and control groups while controlling for baseline scores. Estimated marginal means and SE (adjusted for covariate baselines) are in Table 2.

CRT- a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline CRT scores and conditions ($F(1144) = 1.21, p = .27$). Levene's test indicated that the assumptions of homogeneity of variances was violated ($F(1146) = 26.01, p < .001$). However, the critical value when comparing Hartley's variance ratio ($F_{max} = 2.29$) indicated no reasons for concern. The covariate, baseline CRT score, was significantly related to the post-test CRT score ($F(1145) = 25.64, p < .001, \eta^2 = 0.15$). However, there was a significant effect of condition on the post-test CRT score after controlling for the effect of baseline CRT score ($F(1145) = 60.60, p < .001, \eta^2 = 0.29$). Comparing the estimated marginal means (adjusted for covariate) revealed that participants in the intervention group scored higher on CRT during the post-test than the control group (Table 2). The post-hoc comparison within the intervention group (*t*-test) revealed a significant increase in CRT score $t(79) = 5.31, p < 0.001, d = 1.89$.

CRT Confidence - a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline CRT confidence scores and conditions ($F(1144) = 0.05, p = .82$). Levene's test indicated no issues with the assumptions of homogeneity of variances ($F(1146) = 2.10, p = .15$). The covariate, baseline CRT confidence score, was not significantly related to the post-test CRT confidence score ($F(1145) = 2.69, p = .10, \eta^2 = 0.02$). However, there was a significant effect of condition on the post-test CRT confidence score after controlling for the effect of baseline CRT confidence score ($F(1145) = 84.48, p < .001, \eta^2 = 0.37$) Comparing the estimated marginal means (adjusted for covariate) revealed that participants in the intervention group scored higher on CRT confidence during the post-test than the control group (Table 2). The post-hoc comparison within the intervention group (*t*-test) revealed a significant increase in CRT confidence score (pre-post) $t(79) = 3.30, p = 0.001, d = 13.63$.

Metacognitive sensitivity - a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline Metacognitive sensitivity scores and conditions ($F(1144) = 0.34, p = .56$). Levene's test indicated issues with the assumptions of homogeneity of variances ($F(1146) = 13.36, p < .001$). However, the critical value when comparing Hartley's variance ratio ($F_{max} = 2.25$) indicated no reasons for concern. The covariate, baseline Metacognitive sensitivity score, was not significantly related to the post-test Metacognitive sensitivity score ($F(1145) = 0.02, p = .88, \eta^2 = 0.001$). However, there was a significant effect of condition on the post-test Metacognitive sensitivity score after controlling for the effect of baseline Metacognitive sensitivity score (F

Table 1
Means (Standard Deviations) of CT measures as a function of intervention and time conditions.

Variables	Experiment				Control			
	Pre		Post		Pre		Post	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CRT	3.22	1.50	4.35	1.90	3.36	1.68	2.52	1.11
CRT (Conf)	55.26	11.20	60.30	9.05	49.70	14.82	44.42	10.86
MetaCog_Sensitivity	0.19	0.15	0.23	0.13	0.18	0.12	0.13	0.08
CriTT (Misc.)	21.92	5.62	19.16	4.47	21.75	6.51	23.97	4.35
CriTT (Val)	46.82	8.65	48.34	8.00	51.51	6.35	48.10	5.42
CriTT (Conf)	118.85	20.15	131.75	18.70	120.57	20.33	118.12	16.38
AET (BDR)	0.27	0.28	0.16	0.29	0.23	0.31	−0.12	0.27
AET (ADR)	0.14	0.30	0.24	0.28	0.14	0.15	−0.05	0.16
ATS (Analytic)	55.84	24.01	53.97	24.12	52.94	23.86	47.43	27.31

Table 2
Estimated marginal Means (Standard Errors) and Confidence interval (CI) of post-CT measures between groups when controlling for baseline (co-variate) scores).

Variables	Experiment				Control			
	Post score		CI		Post score		CI	
	Mean	SE	Lower	Upper	Mean	SE	Lower	Upper
CRT*	4.37	0.16	1.41	2.37	2.48	0.18	2.13	2.83
CRT (Conf)*	60.04	1.11	57.83	62.24	44.74	1.21	42.34	47.13
MetaCog_Sensitivity*	0.23	0.01	0.21	0.26	0.13	0.01	0.10	0.16
CriTT (Misc) *	19.16	0.50	18.18	20.14	23.97	0.54	22.91	25.03
CriTT (Val)	48.52	0.75	47.04	50.02	47.88	0.82	46.25	49.50
CriTT (Conf)*	131.84	1.97	127.95	135.73	118.01	2.13	113.79	122.22
AET (BDR)*	0.16	0.03	0.10	0.23	−0.13	0.03	−0.19	−0.06
AET (ADR)*	0.25	0.03	0.20	0.30	−0.05	0.03	−0.10	0.01
ATS (Analytic)	53.86	2.87	48.19	59.54	47.56	3.11	41.40	53.71

Note- all * denotes significant effect of intervention at $p<.001$.

(1145) = 29.32, $p < .001$, $\eta^2 = 0.17$). Comparing the estimated marginal means (adjusted for covariate) revealed that participants in the intervention group scored higher on Metacognitive sensitivity during the post-test than the control group. The post-hoc comparison within the intervention group (*t*-test) revealed a significant increase in Metacognitive sensitivity score (pre-post) $t(79) = 3.36$, $p = 0.001$, $d = 0.29$.

CriTT Misconceptions - a scatterplot indicated linearity and the regression slopes were similar but suggesting a small interaction between baseline CriTT Misconceptions scores and conditions ($F(1144) = 4.26$, $p = .04$). However, Levene's test indicated no issues with the assumptions of homogeneity of variances ($F(1146) = 0.12$, $p = .72$). The covariate, baseline CriTT Misconceptions score, was not significantly related to the post-test CriTT Misconceptions score ($F(1145) = 0.03$, $p = .86$, $\eta^2 = 0.001$). However, there was a significant effect of condition on the post-test CriTT Misconceptions score after controlling for the effect of baseline CriTT Misconceptions score ($F(1145) = 43.19$, $p < .001$, $\eta^2 = 0.23$) Comparing the estimated marginal means (adjusted for covariate) revealed that participants in the intervention group scored lower on CriTT Misconceptions during the post-test than the control group (Table 2). The post-hoc comparison within the intervention group (*t*-test) revealed a significant decrease in CriTT Misconception score (pre-post) $t(79) = 4.47$, $p = 0.002$, $d = 7.70$.

CriTT Valuing - a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline CriTT Valuing scores and conditions ($F(1144) = 0.06$, $p = .81$). Levene's test indicated issues with the assumptions of homogeneity of variances ($F(1146) = 4.38$, $p = .04$). However, the critical value when comparing Hartley's variance ratio ($F_{max} = 1.37$) indicated no reasons for concern. The covariate, baseline CriTT Valuing score, was not significantly related to the post-test CriTT Valuing score (F(1145) = 1.54, $p = .22$, $\eta^2 = 0.01$). The effect of condition on the post-test CriTT Valuing score after controlling for the effect of baseline CriTT Valuing score was also not significant ($F(1145) = 0.32$, $p = .57$, $\eta^2 = 0.002$). The post-hoc comparison within the intervention group (*t*-test) revealed no significant changes in CriTT Valuing score (pre-post) $t(79) = 1.25$, $p = 0.21$.

CriTT Confidence - a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline CriTT Confidence scores and conditions ($F(1144) = 0.55$, $p = .46$). Levene's test indicated no issues with the assumptions of homogeneity of variances ($F(1146) = 0.07$, $p = .79$). The covariate, baseline CriTT Confidence score, was not significantly related to the post-test CriTT Confidence score ($F(1145) = 2.55$, $p = .11$, $\eta^2 = 0.01$). However, there was a significant effect of condition on the post-test CriTT Confidence score after controlling for the effect of baseline CriTT Confidence score ($F(1145) = 22.69$, $p < .001$, $\eta^2 = 0.13$) Comparing the estimated marginal means (adjusted for covariate) revealed that participants in the intervention group scored higher on CriTT Confidence during the post-test than the control group (Table 2). The post-hoc comparison within the intervention group (*t*-test) revealed a significant increase in CriTT confidence score (pre-post) $t(79) = 4.62$, $p < 0.001$, $d = 24.95$.

AET-BDR - a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline AET-BDR scores and conditions ($F(1144) = 0.11$, $p = .74$). Levene's test indicated that the assumptions of homogeneity of variances was violated ($F(1146) = 17.73$, $p < .001$). However, the critical value when comparing Hartley's variance ratio (Fmax = 0.88) indicated no reasons for concern. The covariate, baseline AET-BDR score, was not significantly related to the post-test AET-BDR score ($F(1145) = 0.11$, $p = .74$, $\eta^2 = 0.02$). However, there was a significant effect of condition on the post-test AET-BDR score after controlling for the effect of baseline AET-BDR score ($F(1145) = 37.04$, $p < .001$, $\eta^2 = 0.20$). Comparing the estimated marginal means (adjusted for covariate) revealed that participants in both conditions scored lower on AET-BDR post-test (Table 2). The post-hoc comparison within the intervention group (*t*-test) revealed a significant decrease in AET-BDR scores (pre-post) $t(79) = -2.33$, $p = 0.02$, $d = 0.41$.

AET-ADR - a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline AET-BDR scores and conditions ($F(1144) = 2.10$, $p = .15$). Levene's test indicated no issues with the assumptions of homogeneity of variances ($F(1146) = 0.52$, $p = .47$). The covariate, baseline AET-ADR score, was not significantly related to the post-test AET-ADR score ($F(1145) = 3.40$, $p = .07$, $\eta^2 = 0.02$). However, there was a significant effect of condition on the post-test AET-ADR score after controlling for the effect of baseline AET-ADR score ($F(1145) = 58.49$, $p < .001$, $\eta^2 = 0.29$). Comparing the estimated marginal means (adjusted for covariate) revealed that participants in the intervention group scored higher on AET-ADR post-test during the post-test than the control group (Table 2). The post-hoc comparison within the intervention group (*t*-test) revealed a significant increase in AET-ADR scores (pre-post) $t(79) = 2.03$, $p = 0.04$, $d = 0.44$.

ATS - a scatterplot indicated linearity and the regression slopes were similar suggesting no interaction between baseline ATS scores and conditions ($F(1144) = 0.47$, $p = .49$). Levene's test indicated no issues with the assumptions of homogeneity of variances ($F(1146) = 2.03$, $p = .16$). The covariate, baseline ATS score, was not significantly related to the post-test ATS score ($F(1145) = 0.44$, $p = .22$, $\eta^2 = 0.006$). The effect of condition on the post-test ATS score after controlling for the effect of baseline ATS score was also not significant ($F(1145) = 2.21$, $p = .14$, $\eta^2 = 0.01$). The post-hoc comparison within the intervention group (*t*-test) revealed no significant changes in ATS score (pre-post) $t(79) = 0.53$, $p = 0.59$.

4. Discussion

The current study examined the effect of CT intervention on the selected measures of CT ability, attitudes, and behaviours. As hypothesised, results revealed a significant post-intervention increase for all CT measures (excluding misconceptions and belief driven responses which were predicted to be reduced) except analytic writing style and CriTT valuing. The current study replicated previous findings (Simonovic et al., 2022) that a brief CT intervention can increase abilities, attitudes, and confidence related to CT. Students' confidence in CT was higher in the intervention group after the intervention, while avoidance or misconceptions of CT decreased after the intervention. Compared to the control group, students in the intervention group scored higher on measures of cognitive reflection and confidence in cognitive reflection after the intervention. Moreover, compared to the control group, students in the intervention group revealed a significant decrease in belief-driven responding and a significant increase in argument-driven responding. These findings are now considered in detail.

Intervention effects. The results indicate that our teaching intervention, underpinned by dual-process and meta-reasoning theories was a highly effective method to enhance CT skills at least in the short term. This is consistent with research that CT can be enhanced by mixed approach instruction (Abrami et al., 2008, 2015; Tiruneh et al., 2014). These data further indicate that reflective thinking, self-regulation, beliefs, and attitudes represent important factors for enhancing CT skills (Cloete, 2018; Dwyer et al., 2014; Dwyer & Walsh, 2020; Stupple et al., 2017a). This is also in line with previous research (Simonovic et al., 2022) and with the suggestion that Type 2 processing such as conflict detection and explicit resolution may help students to become aware of the need for overriding previously held beliefs and help them sustain cognitive reflection to deploy appropriate mindware (knowledge required for normative reasoning and judgment) for explicit resolution (Stanovich 2009; De Neys & Bonnefon, 2013; Bonnefon, 2016). Thus, Type 2 processes may be a prerequisite for CT and focusing interventions on specific Type 2 processes can be a good way of facilitating CT skills.

The score on the measure of beliefs and attitudes of CT (CriTT) improved in the intervention group. Confidence in CT significantly increased from pre- to post-testing. Avoidance or misconception of CT was decreased only in the intervention group. This corresponds with Simonovic et al.'s (2022) previous results and demonstrates that the effect of the intervention generalises to a general population of UK higher education students (rather than specifically supporting online psychology students). It further indicates that CT attitudes and beliefs are malleable in the presence of a “how to think” intervention. This effect on students' confidence to demonstrate and develop their CT skills may further enhance CT performance (Ennis. 1985). On this basis, we argue that when developing CT interventions, including consideration of student beliefs and how they formulate their belief system (Lamont, 2020) should be given consideration due to the connections between students' beliefs and attitudes and CT performance (Stupple et al., 2017a). It is important to note that we have not demonstrated that these benefits persist in the long term or that the effect generalises beyond the set of standardised thinking and reasoning tasks used in this study, however these limitations will be explored in future studies.

An important element of Type 2 processes is sustained inhibition whereby the individual must inhibit an initial, intuitive response and engage in more reflective thinking to reach a deliberate judgement based on the relevant criteria (Bonnefon, 2018; Evans & Stanovich 2013). In line with previous research, our data support a theoretically and practically important relationship between cognitive reflection and CT skills (Kember et al., 2000; Simonovic et al., 2022; Stupple et al., 2017a). Scores on CRT were higher after the intervention which is line with recent study whereby a brief intervention improved cognitive reflection performance (Calvillo et all., 2022). The intervention in the present study is likely to have enhanced performance by encouraging participants to slow down and consider alternative responses when engaging in challenging thinking tasks. It has been established previously that the CRT measure correlates highly with SAT scores (the measure of readiness for college-level education), academic achievement and student's GPA (Frederick 2005; Simonovic et al., 2022; Stupple et al., 2017a). It could be argued that the CRT measures students' ability to engage executive functions to sustain inhibition of tempting incorrect responses and to effortfully monitor and self-regulate their thinking in search of another possible response. As such the CRT, maps on the CT framework that emphasises, self-regulation, reflective judgment, and cognitive inhibition as important facets of CT skills (Dwyer et al., 2014) and may represent an effective tool in understanding CT in higher education that may be useful in identifying students in need of support in developing CT during their studies. This finding is, moreover of great interest to the dual process theory debate over the CRT which contrasts the importance of cognitive miserliness versus the possession of the requisite mindware to solve the problems – these data indicate that enhancing participant mindware in a brief intervention can significantly improve performance, undermining the predominant cognitive miserliness account (see e.g., Stanovich, 2018; Stupple et al. 2017b)

Several aspects of metacognition have been identified as important for the development and execution of CT skills (e.g., task monitoring, self-regulation) with analysis, evaluation, and inferences as key elements to CT (See Dwyer et al., 2014 for a review). Relatively few studies have directly examined the relationship between CT and metacognition (e.g., Ku & Ho, 2010; Magno, 2010; Bensley & Spero, 2014). Our findings indicate that a short CT intervention can help students to become better calibrated in judging their reflective reasoning performance. This is in line with an argument that closing the mindware gap depends on knowledge, skills, and accuracy (Bensley & Spero, 2014). This finding is again consistent with the argument from Stupple et al. (e.g., 2017b) that the CRT is not a good measure of cognitive miserliness for participants who lack the relevant mindware as the improved mindware enhanced both performance and awareness of performance. A lack of awareness of the need for deliberative or critical thinking cannot be equated with cognitive miserliness or an inability to think critically. These data indicate that teaching CT skills should take a comprehensive “how to think” that encourages the metacognitive monitoring of accuracy and enhances the accuracy of that monitoring and also by encouraging the development of appropriate mindware.

The intervention group also outperformed the control group on the Argument Evaluation Test (AET). Participants in the intervention group scored lower on belief-driven responding (BDR; indicative of biased Type 1 processing) but higher on argument-driven responding (ADR: indicative of reflective Type 2 processing) compared to the control group. The AET measures two important aspects of argumentation that are essential for CT (Stanovich & West, 1997) and are related to analysis, evaluation, and inference of evidence. Notably, at the core of Facione's (1990) definition and Dwyer's et al. (2014) proposed framework, the ability to recognize the structure of an argument, evaluation of the credibility of claims and arguments and drawing a reasonable conclusion are important for the development of CT skills and application of CT skills in real-life situations.

We further contend that teaching strong argumentation could be an important factor in closing the mindware gap, where a student does not have access to adequate knowledge to synthesize a correct response that will override incorrect belief driven responses (Stanovich, 2009, 2016). As such, one of the strategies for CT interventions would be to support students in developing the mindware required for an explicit resolution to provide them with CT tools (reflective, evaluative, argumentative, and logical) to help them avoid cognitive bias and heuristic thinking and reach reasonable and judicious conclusions.

The effect of the intervention on analytic writing between the conditions was not significant. Tausczik & Pennebaker (2010) argued that thinking varies in depth and complexity and that depth is expressed in the words people use. The use of cognitive complexity and cognitive or analytic words can reflect the richness of reasoning the writer has engaged in Pennebaker et al. (2014). Our focus in this study was on the use of analytic words because they are particularly relevant for CT (Pennebaker et al., 2003). Our results could be explained by the topic choice of the essay. We asked the students to describe an accomplishment or a captivating topic that could have been perceived as emotionally leaden and although Pennebaker et al. (2003) argued that the incidence of analytic words should increase regardless of individual emotional upheaval, we however did not find this to be the case. A writing task more reliant on argument and analysis may produce a different outcome and warrants further research.

Limitations and future suggestions. In our previous study, CT intervention improved student's academic achievement (measured by GPA), however, in this study, academic achievement was not measured due to the diversity of our sample coming from different Universities. Future studies should consider measuring academic achievement but also other aspects beyond academic achievements such as real-life events and accomplishments (Butler et al., 2017). Second, we did not account for motivation. There is a possibility that the students self-selected to take part in the study because they are motivated to learn critical thinking and perceive it as an essential skill in their studies. Although the waiting list control group addresses some of these concerns, we argue that future studies should account for motivation as a potential confounding variable. (e.g., Rigs & Hellyer-Riggs, 2014). Third, the significant interactions in this study indicate that at least some transfer of knowledge took place, but our study was conducted in a limited time frame, and it is not certain if the intervention would have lasting effects. Although carry-over effects from pre to post-test cannot be fully dismissed, these would apply to both intervention and control groups. Fourth, there are further limitations with the use of the CRT. Some items are increasingly well known, and most standard CRT questions are mathematical (e.g., Campitelli & Gerrans, 2014). This was mitigated by including the verbal version of the CRT, to reduce reliance on mathematical ability with careful item selection and rigorous counterbalancing. Future studies that are focused on LIWC analyses should use formally established topics with a focus on CT and require analysis, argument, and reflection to tease out possible interactions between analytical writing style and CT – the nature of the present writing task may not have facilitated this.

Finally, our approach to metacognition was based upon the meta-reasoning framework proposed by Ackerman and Thompson (2017) and used post answer confidence judgements to examine metacognitive sensitivity, this requires further nuanced research if we are to make broader claims about metacognition and CT. For example, the increased metacognitive sensitivity observed here could be a result of participants developing a clearer understanding of the task requirements rather than enhancing metacognition per se. It would be prudent for future research to include a more comprehensive set of metacognitive measures to unpack the relative roles of metacognitive awareness, metacognitive knowledge, and metacognitive skills in critical thinking interventions (see e.g., Schraw, 1998).

Our intervention can be used in an educational context to help students understand the importance of CT and to encourage students to engage in new ways to think and focus in a more reflective, deliberate, and analytic way. Type 2 processing is necessary but not sufficient for CT because it can magnify biases through, for example, reasoning motivated by partisan beliefs (e.g., Van Bavel & Pereira, 2018). CT nonetheless entails higher-order thinking and reasoning that can be improved by training (Stanovich & Stanovich, 2010, 2016). Thus, targeting Type 2 processes with intervention may enhance attitudes and beliefs about CT and facilitate strong cognitive reflections, evaluation and inference of arguments that can help students to improve their performance on the most challenging of academic tasks.

This is one of the first studies to examine the effect of an explicit mixed-method intervention on such a comprehensive set of measures of CT skills and discussion motivated by a dual process theory of thinking and meta-reasoning framework. The results indicate that CT can be significantly enhanced with a brief online intervention, and targeting specific aspects of Type 2 processes such as cognitive reflection, evaluation of arguments and drawing reasonable conclusions can enhance reflective thinking and argumentation and impact attitudes and beliefs about CT.

Author's contributions

All author(s) contributed to the conceptualization, design and writing of the paper.

Funding

This work was supported by internal funding from University of Derby.

Declaration of Competing Interest

The author(s) declare no potential competing interests with respect to the research, authorship, and/or publication of this article.

Data availability

Data are available on request from B. Simonovic.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tsc.2023.101362.

References

Abe, J. A. A. (2020). Big five, linguistic styles, and successful online learning. *Internet and Higher Education, 45*(January), Article 100724. https://doi.org/10.1016/j.iheduc.2019.100724

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

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., et al. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102–1134. https://doi.org/10.3102/0034654308326084

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General, 143*(3), 1349–1368.

Ackerman, R., Parush, A., Nassar, F., & Shtub, A. (2016). Metacognition and system usability: Incorporating metacognitive research paradigm into usability testing. *Computers in Human Behavior, 54*, 101–113. https://doi.org/10.1016/j.chb.2015.07.041

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences, 21*(8), 607–617. https://doi.org/10.1016/j.tics.2017.05.004

Bailin, S., & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (Eds.), *The blackwell guide to the philosophy of education. Oxford* (pp. 181–193). Blackwell.

Barrie, S. (2006). Understanding what we mean by the generic attributes of graduates. *Higher Education, 51*, 215–241.

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

Bonnefon, J. F. (2018). The pros and cons of identifying critical thinking with system 2 processing. *Topoi, 37*, 113–119.

Boyd, R. L., & Pennebaker, J. W. (2015). A way with words: Using language for psychological science in the modern era. In C. V. Dimofte, C. P. Haugtvedt, & R. F. Yalch (Eds.), *Consumer psychology in a social media world* (pp. 222–236). Routledge. https://doi.org/10.4324/9781315714790.

Butler, H. A., Pentoney, C., & Bong, M. P. (2017). Predicting real-world outcomes: Critical thinking ability is a better predictor of life decisions than intelligence. *Thinking Skills and Creativity, 25*, 38–46. https://doi.org/10.1016/j.tsc.2017.06.005

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307.

Calvillo, D. P., Bratton, J., Velazquez, V., Smelter, T. J., & Crum, D. (2022). Elaborative feedback and instruction improve cognitive reflection but do not transfer to related tasks. *Thinking & Reasoning*. https://doi.org/10.1080/13546783.2022.2075035

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modelling approach. *Memory & Cognition, 42*(3), 434–447. https://doi.org/10.3758/s13421-013-0367-9

Carroll, D. W. (2007). Patterns of student writing in a critical thinking course: A quantitative analysis. *Assessing Writing, 12*(3), 213–227.

Cloete, M. (2018). The impact of an integrated assessment on the critical thinking skills of first-year university students. *Accounting Education, 27*(5), 479–494. https://doi.org/10.1080/09639284.2018.1501717

Davies, T. (2006). Creative teaching and learning in Europe: Promoting a new paradigm. *The Curriculum Journal, 17*(1), 37–57.

De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends Cognitive Science, 17*, 172–178.

Duro, E., Elander, J., Maratos, F. A., Stupple, E. J. N., & Aubeeluck, A. (2013). In search of critical thinking in psychology: An exploration of student and lecturer understandings in higher education. *Psychology Learning & Teaching, 12*(3), 275–281. https://doi.org/10.2304/plat.2013.12.3.275

Dwyer, C., & Walsh, A. (2020). An exploratory quantitative case study of critical thinking development through adult distance learning. *Educational Technology Research and Development, 68*(1), 17–35. https://doi.org/10.1007/s11423-019-09659-2

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning, 7*, 219–244.

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

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 44. Retrieved from http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_198510_ennis.pdf.

Ennis, R. H. (2003). Critical thinking assessment. In D. Fasko (Ed.), *Critical thinking and reasoning* (pp. 293–310). Cresskill, NJ: Hampton Press.

Ennis, R. H. (2016). Critical thinking across the curriculum: A vision. *Topoi, 37*(1), 1–20. https://doi.org/10.1007/s11245-016-9401-4

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives in Psychological Science, 8*, 223–241. https://doi.org/10.1177/1745691612460685

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: California Academic Press (ERIC document ED 315-423).

Facione, P. A., & Facione, N. C. (2001). Analyzing explanations for seemingly irrational choices: Linking argument analysis and cognitive science. *International Journal of Applied Philosophy, 15*(2), 267–286. Retrieved from https://insightassessment.com/wp-content/uploads/ia/pdf/IJAP_Analysis_Paper.pdf.

Ferguson, L. E., & Bråten, I. (2013). Student profiles of knowledge and epistemic beliefs: Changes and relations to multiple-text comprehension. *Learning and Instruction, 25*, 49–61.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42. https://doi.org/10.1257/089533005775196732

Gambrill, E. (2006). Evidence-based practice and policy: Choices ahead. *Research on Social Work Practice, 16*(3), 338–357. https://doi.org/10.1177/1049731505284205

Gangemi, A., Bourgeois-Gironde, s., & Mancini, F. (2015). Feelings of error in reasoning—in search of a phenomenon. *Thinking & Reasoning, 21*(4), 383–396.

Ghanizadeh, A. (2017). The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *Higher Education, 74*(1), 101–114. https://doi.org/10.1007/s10734-016-0031-y

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking (4th ed.)*. New Jersey: Laurence Erlbaum Associates, 10.1002/(SICI)1099-0720(199704)11:2≤179::AID-ACP453≥3.0.CO;2-S.

Halpern, D. F. (2010). *The Halpern critical thinking assessment: Manual*. Vienna: Schuhfried.

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

Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. *Review of Educational Research, 86*(2), 431–468. https://doi.org/10.3102/0034654315605917

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341–350.

Kember, D., Leung, D., Jones, A., & Loke, A. Y. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment and Evaluation in Higher Education, 25*(4), 380–395. https://doi.org/10.1080/713611442

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490. https://psycnet.apa.org/doi/10.1037/0033-295X.103.3.490.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121. https://doi.org/10.1037/0022-3514.77.6.1121

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*(1), 70–76. https://doi.org/10.1016/j.tsc.2009.02.001

Ku, K. Y., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning, 5*, 251–267.

Lamont, P. (2020). The construction of "critical thinking": Between how we think and what we believe. *History of Psychology*. https://doi.org/10.1037/hop0000145. Advance online publication.

Magno, C. (2010). The role of metacognitive skills in developing critical thinking. *Metacognition and Learning, 5*(2), 137–156. https://doi.org/10.1007/s11409-010-9054-4

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In *In the psychology of learning and motivation: Advances in research and theory, 26* pp. 125–173). New York: Academic. ed. GH Bower.

Osborne, R., Dunne, E., & Farrand, P. (2013). Integrating technologies into 'authentic' assessment design: An affordance approach. *Research in Learning Technology, 21*, 1–18. https://doi.org/10.3402/rlt.v21i0.21986

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PloS One, 9*(12), Article e115844.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language as an individual differences. *Journal of Personality and Social Psychology, 77*, 1296–1312.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577.

Puig, B., Blanco-Anaya, P., Bargiela, I. M., & Crujeiras-Pérez, B. (2019). A systematic review on critical thinking intervention studies in higher education across professional fields. *Studies in Higher Education, 44*(5), 860–869. https://doi.org/10.1080/03075079.2019.1586333

Rigs, L. W., & Hellyer-Riggs, S. (2014). Development and motivation in/for critical thinking. *Journal of College Teaching & Learning, 11*(1), 1–8.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*(1–2), 113–125. https://psycnet.apa.org/doi/10.1023/A:1003044231033.

Simonovic, B., Vione, K. C., Fido, D., Stupple, E. J. N., Martin, J., & Clarke, R. (2022). The impact of attitudes, beliefs, and cognitive reflection on the development of critical thinking skills in online students. *Online Learning, 26*(2), 1–21.

Stanovich, K. E. (2016). The comprehensive assessment of rational thinking. *Educational Psychologist, 51*(1), 23–34. https://doi.org/10.1080/00461520.2015.1125787

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking and Reasoning, 24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*(2).

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672–695. https://doi.org/10.1037/0022-3514.94.4.672

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. In JSBT Evans, & K Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford: Oxford University Press.

Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. In D Preiss, & RJ Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching and human development* (pp. 195–237). New York: Springer.

Stupple, E. J. N, Maratos, F. A., Elander, J., Hunt, T. E., Cheung, K. Y. F., & Aubeeluck, A. V (2017a). Development of the Critical Thinking Toolkit (CriTT): A measure of student attitudes and beliefs about critical thinking. *Thinking Skills and Creativity, 23*, 91–100. https://doi.org/10.1016/j.tsc.2016.11.007

Stupple, E. J. N., Pitchford, M., Ball, L. J., Hunt, T. E., & Steel, R. (2017b). Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the cognitive reflection test. *PloS One, 12*(11). https://doi.org/10.1371/journal.pone.0186404

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning, 20*(2), 215–244.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Bal, l L. J., Brack, H., Ophir, Y., et al. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making, 11*(1), 99.

Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies, 4*(1). https://doi.org/10.5539/hes.v4n1p1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Tsui, L. (2002). Fostering critical thinking through effective pedagogy. *Research in Higher Education, 40*(2), 740–763. https://doi.org/10.1080/00221546.2002.11777179

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences, 22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Vertecchi, B., Poce, A., Agrusti, F., & Re, M. R. (2017). Pen or keyboard. An empirical study on the effects of technology on writing skills. *Cadmo, Giornale italiano di Pedagogia Sperimental, 25*(1), 33–44. https://doi.org/10.3280/cad2016-002004

Watson, G., & Glaser, E. M. (1964). *Watson-Glaser critical thinking appraisal manual: Forms Ym and Zm*. New York: Harcourt, Brace, and World.