



Identifying characteristics associated with higher education teachers' Cognitive Reflection Test performance and their attitudes towards teaching critical thinking

Eva M. Janssen^{a, *}, Wietse Meulendijks^a, Tim Mainhard^a, Peter P.J.L. Verkoeijen^{b, c}, Anita E.G. Heijltjes^c, Lara M. van Peppen^b, Tamara van Gog^a

^a Department of Education, Utrecht University, Heidelberglaan 1, 3584, CS, Utrecht, the Netherlands

^b Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062, PA, Rotterdam, the Netherlands

^c Learning and Innovation Center, Avans University of Applied Sciences, Hogeschoollaan 1, 4818, CR, Breda, the Netherlands

HIGHLIGHTS

- Teachers' disposition toward effortful thinking was positively related to their Cognitive Reflection Test (CRT) performance.
- Teaching in a more technological domain was positively related to teachers' CRT performance.
- Teachers' level of education was positively related to their CRT performance.
- Dispositions toward effortful and open-minded thinking were positively related to perceived relevance of teaching CT.
- Teachers' confidence in CRT performance rather than actual performance was related to perceived competence in teaching CT.

ARTICLE INFO

Article history:

Received 1 October 2018
Received in revised form
7 May 2019
Accepted 24 May 2019
Available online 7 June 2019

Keywords:

Critical thinking
CRT
Teaching attitudes
Thinking dispositions
Teaching and teacher education
Higher education

ABSTRACT

The aim of this study was to identify characteristics that are related to higher education teachers' ($N = 263$) Cognitive Reflection Test (CRT) performance, which assesses an important aspect of critical thinking (CT), and their attitudes towards teaching CT more generally. Results of a structural equation model showed that a stronger disposition towards effortful thinking, teaching in a more technological domain, and a higher level of education were related to a better CRT performance. Thinking dispositions were also related to teachers' perceived relevance of teaching CT. Confidence in CRT performance rather than actual performance was related to perceived competence in teaching CT.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

The aim of this study was to identify characteristics that are related to higher education teachers' Cognitive Reflection Test (CRT) performance, which assesses an important aspect of critical thinking

(CT), and their attitudes towards teaching CT more generally. Teaching CT is an important topic in higher education, since one of the major ambitions of higher education is to foster students' CT-skills (National Research Council, 2012). Teachers have an important role to play in this process, as it has been shown that students' CT-skills do not develop automatically as a by-product of higher education (Arum & Roksa, 2011; Pascarella, Blaich, Martin, & Hanson, 2011) and students need explicit instruction to improve their CT-skills (Abrami et al., 2015; Heijltjes et al., 2014). Remarkably, even though reviews on teaching CT highlight the crucial role of the teacher (Abrami et al., 2008, 2015; Pithers & Soden, 2000; Ritchhart & Perkins, 2005), studies on teachers' CT are scarce and mostly

* Corresponding author.

E-mail addresses: e.m.janssen@uu.nl (E.M. Janssen), wietsemeulendijks@gmail.com (W. Meulendijks), m.t.mainhard@uu.nl (T. Mainhard), p.p.j.l.verkoeijen@essb.eur.nl, ppjl.verkoeijen@avans.nl (P.P.J.L. Verkoeijen), aeg.heijltjes@avans.nl (A.E.G. Heijltjes), vanpeppen@essb.eur.nl (L.M. van Peppen), tvangog@uu.nl (T. van Gog).

focused on pre-service teachers. According to the few available studies, higher education teachers (i.e., postsecondary teachers) may not have a concrete understanding of what CT encompasses and how they can teach it (Choy & Cheah, 2009; Stedman & Adams, 2012). This would be problematic for teaching CT, because two basic requirements for being able to teach a particular subject or skill are possessing the skill oneself (Hattie, 2003; Jones & Moreland, 2003) and having a positive attitude towards teaching it (Klassen & Tze, 2014; Van Aalderen-Smeets & Walma van der Molen, 2013; 2015). Variables that play a role in higher education teachers' CT-skills and attitudes towards teaching CT have not yet been identified, yet knowledge on these variables can be informative for research on how to better equip teachers for teaching CT. Thus, as a first step, the present study investigated what teacher characteristics were associated with higher education teachers' CRT performance (as an important aspect of CT) and with positive attitudes towards teaching CT more generally.

1.1. Critical thinking and the Cognitive Reflection Test (CRT)

An essential aspect of CT is to avoid bias in reasoning and decision-making (i.e., rational thinking). Bias is said to occur when a reasoning process results in a systematic deviation from a norm when choosing actions or estimating probabilities (Stanovich, West, & Toplak, 2016; Tversky & Kahneman, 1974). Examples of biases that commonly occur are that people tend to make predictions based on their intuition without taking the probability of an outcome into account (Kahneman & Tversky, 1973), tend to look for confirmation instead of falsification when testing hypotheses (Wason, 1968), tend to infer conclusions based on personal beliefs in violation of logic (Evans, Handley, & Harper, 2001), and tend to make choices that are affected by irrelevant contextual information (Jacowitz & Kahneman, 1995; Tversky & Kahneman, 1986). Biases in reasoning can have serious consequences for decision-making in both daily life and complex professional environments (Heijltjes et al., 2014; Lunn, 2013; Thompson & Schumann, 1987; Toplak, West, & Stanovich, 2017). Researchers have been studying bias empirically with reasoning problems in which an intuitively cued heuristic response conflicts with elementary logical principles (also called heuristics-and-biases tasks). One test that has been studied extensively in the reasoning and decision literature is Frederick's (2005) 3-item Cognitive Reflection Test (CRT). The most famous problem in this test is:

A bat and a ball together cost \$1.10. The bat costs \$1 more than the ball. How much does the ball cost?

Most reasoners intuitively conclude that the ball must cost 10 cents ($\$1 + \$0.10 = \$1.10$). However, this conclusion is incorrect¹ because in this scenario the bat costs 90 cents more than the ball instead of \$1. After some reflection, it should become clear that the

¹ In research on reasoning, it is debated whether heuristic responses should be labelled "incorrect" or "biased" (for a review, see Stanovich & West, 2000). For the sake of simplicity we use the terms "correct" response or "logical" response for the responses that are considered normatively correct following the rules of logic or probability and "incorrect" for responses that are not normatively correct according to the rules of logic or probability.

² The algebraic equation behind the problem is:

$$\begin{aligned}(x + 1) + x &= 1.10. \\ 2x + 1 &= 1.10 \\ 2x &= 0.10 \\ x &= 0.05.\end{aligned}$$

correct answer requires a different calculation leading to the conclusion that the ball costs 5 cents ($\$1.05 + \$0.05 = \$1.10$).² The logical answer of 5 cents does not require strong mathematical skills, yet a number of studies have showed that even educated reasoners fail to solve the problem correctly (Frederick, 2005; Toplak, West, & Stanovich, 2014). CRT items are designed to measure people's tendency to override their intuitive incorrect response and to engage in further reflection that leads to the correct response (Frederick, 2005). The original 3-item CRT and extended versions have been shown to reliably predict a person's ability to make unbiased judgments and decisions in a wide variety of contexts (Pennycook, Cheyne, Koehler, & Fugelsang, 2015; Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2016; Toplak et al., 2014; Toplak, West, & Stanovich, 2011). In this study, we used the 7-item CRT developed by Toplak et al. (2014) to measure an important aspect of CT.

1.1.1. Variables associated with CRT performance

In order to avoid a biased judgment on heuristics-and-biases tasks like for example the bat-and-ball problem, dual process theories explain that one needs to *override* an intuitive/heuristic Type 1 response ($\$1 + \$0.10 = \$1.10$) with a more effortful/logical Type 2 response ($\$1.05 + \$0.05 = \$1.10$; Evans, 2008; Kahneman, 2011; Kahneman & Frederick, 2005). Within the dual-process literature, researchers explain (in)correct performance on such tasks with individual differences in people's thinking dispositions, their available mindware, and their cognitive ability (Heijltjes et al., 2015; Frederick, 2005; Klaczynski, 2014; West, Toplak, & Stanovich, 2008). People with strong rational thinking dispositions – people with the tendency to enjoy and engage in effortful thinking and actively open-minded thinking – are more inclined to detect the need for a Type 1 override than people with less strong rational thinking dispositions (Stanovich, 2011; Stanovich et al., 2016). However, merely detecting the need to override is necessary but not sufficient for a good decision. One also needs to possess the requisite *mindware*, that is, the declarative knowledge and skills needed for the reasoning situation (e.g., in the example above, algebraic mathematical skills), and sufficient working memory capacity to start and sustain an override. All three variables are necessary conditions for being able to successfully perform the heuristics-and-biases tasks. For the CRT, numerous studies showed that thinking dispositions, numeracy mindware, and cognitive ability were independent predictors of successful performance (Campitelli & Gerrans, 2014; Toplak et al., 2011, 2014). However, no research with teachers has been conducted yet. Knowledge on what teacher characteristics are related to a better CRT performance could provide some first insights into the variables that play a role in teachers' CT-skills (required for teaching CT). Based on the studies outlined above, we hypothesized that teachers with a stronger disposition towards effortful and actively open-minded thinking would have a better CRT performance (hypothesis 1a); that teachers from technological domains would perform better than teachers from economical and societal domains respectively, because the required numeracy mindware for the CRT is taught most explicitly in technological domains, followed by economics and society, respectively (hypothesis 1b); and that teachers with a higher level of education – associated with cognitive ability – would perform better on the CRT (hypothesis 1c).

1.2. Teaching attitudes

In addition to the skill of thinking itself, believing that one is competent in teaching CT to students (perceived competence) and believing that teaching these skills is relevant (perceived relevance) may be positive antecedents of effective teaching. More general

motivational theories, such as expectancy-value theory, frame these two factors in terms of expectancy for success (perceived competence) and task value (perceived relevance); both are viewed as direct predictors of task performance and persistence (Eccles & Wigfield, 2002). In line with this framework, teaching research has indeed shown that teachers who have a positive attitude towards the relevance of teaching a subject (i.e., high task value) and confidence in their ability to do so (i.e., high expectancy of success) engage in more effective teaching (Klassen & Tze, 2014; Van Aalderen-Smeets & Walma van der Molen, 2013, 2015). For instance, studies in the domain of science teaching in primary school showed that these beliefs were positively related to joy in teaching science (Van Aalderen-Smeets & Walma van der Molen, 2013) and that an intervention focused on changing teachers' professional attitudes on science education and personal attitudes towards science in general, positively affected both perceived relevance and competence, as well as self-reported science teaching behavior in the classroom (Van Aalderen-Smeets & Walma van der Molen, 2015). The authors suggested that teachers' personal attitudes on science in general positively affect their professional attitudes towards teaching it. Given the importance of teaching attitudes for effective teaching, it is relevant to identify what teachers characteristics are related to positive attitudes towards teaching CT. However, it has not yet been investigated what characteristics are related to teachers' perceived relevance of and perceived competence in teaching CT. Based on findings by Van Aalderen-Smeets and Walma van der Molen (2015) in the domain of science, we hypothesized that teachers with a stronger disposition towards effortful and actively open-minded thinking – as an expression of personal attitude on CT in general – would perceive the teaching of CT as more relevant (hypothesis 2a). Additionally, given that performance attainment is one of the principal sources of expectancy and value (Bandura, 1982; Eccles & Wigfield, 2002), we hypothesized that teachers with a better CRT performance would perceive the teaching of CT as more relevant (hypothesis 2b) and would perceive themselves as more competent in teaching it (hypothesis 3).

1.3. The present study

The aim of the present study was to identify teacher characteristics that play a role in three variables that we considered important for effectively teaching CT: (1) teachers' CRT performance (as important aspect of CT-skills); (2) teachers' perceived relevance of teaching CT (3) and teachers' perceived competence in

teaching CT more generally. Based on the literature and hypotheses outlined above, we constructed one model testing all of our hypotheses (see Fig. 1). In addition, we also explored whether CRT performance mediated the relationship between thinking dispositions and perceived relevance of teaching CT (because we hypothesized that thinking dispositions were positively associated with CRT performance and that, subsequently, CRT performance was positively associated with perceived relevance, see Fig. 1).

1.3.1. Exploratory analyses

Potentially, teachers' calibration with regard to their CRT performance may also be relevant for their teaching attitudes and effective teaching behavior. Calibration reflects the degree to which individuals' judgments about their capability correspond to their actual capability (Lichtenstein, Fischhoff, & Phillips, as cited in Alexander, 2013). A teacher with a realistic estimation of her CT-skills knows what CT-skills she already possesses and what skills need some extra practice before being able to teach them. An unrealistic estimation, however, may negatively affect teaching attitudes and behavior. For instance, a teacher with high ability in a particular CT-skill who judges his ability very low (under-confidence), may feel incompetent in teaching it and subsequently avoid teaching the skill to his students. Vice versa, a teacher with a low ability in a certain CT-skill who judges her ability very high (over-confidence), may feel very competent in teaching it but may teach the skill inadequately to her students. To gain insight in teachers' understanding of their own thinking skills, we used their confidence judgments regarding their CRT performance to explore how well they were calibrated and, additionally, whether these confidence judgments mediated our hypothesized relationship of CRT performance with perceived relevance of and perceived competence in teaching CT.

2. Method

2.1. Participants

Participants were teachers from a Dutch university of applied sciences. All 1378 teachers of this university received a request to participate via email. A total of 319 teachers started the questionnaire after providing informed consent. This response rate of 23.2% was about one standard deviation below the average of response rate of surveys in educational sectors (for a review, see Baruch & Holtom, 2008), but could be expected because the announced survey was relatively long (i.e., 30 min), which negatively affects

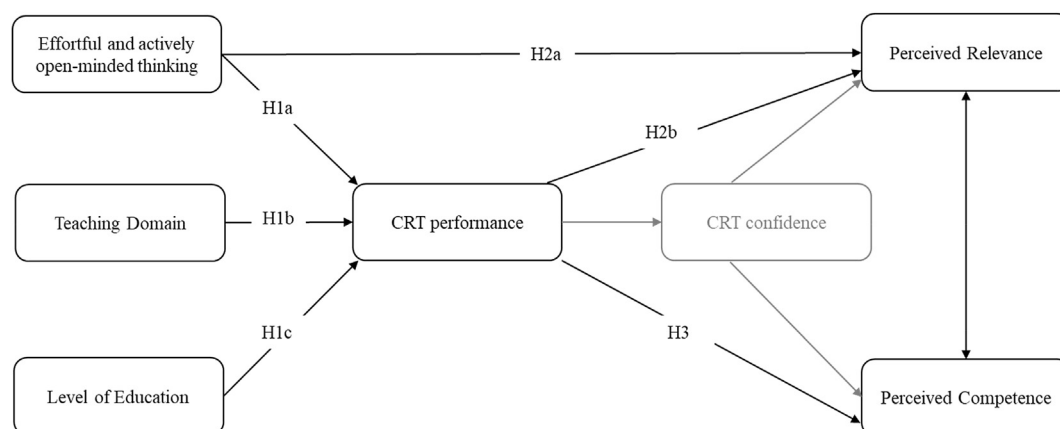


Fig. 1. Hypothesized relationships between teachers' thinking dispositions, teaching domain, level of education, CRT performance, CRT confidence, Perceived Relevance of, and Perceived Competence. Grey indicates exploratory analyses. CRT = Cognitive Reflection Test.

response rates (Galesic & Bosnjak, 2009; Marcus, Bosnjak, Lindner, Pilischenko, & Schütz, 2007). Additionally, we had to exclude data of 56 teachers: four because of noncompliance (e.g., answering with the same response to each question) and 52 because they already dropped out before completing the demographic questions or the CRT (i.e., the first task). Thus, self-selection bias seemed to play a role here; we return to this issue in the discussion. The final sample included 263 teachers (41.4% female; age: $M = 46.3$ years, $SD = 10.7$; teaching experience: $M = 9.8$ years, $SD = 8.7$).

2.2. Materials, procedure, and data analysis

We used an online survey with a forced response-format generated using Qualtrics Survey Software (Qualtrics, Provo, UT; <http://www.qualtrics.com>). The survey addressed four topics in a fixed order: (1) demographics, (2) CRT, (3) teaching attitudes, and (4) thinking dispositions. The survey was in Dutch for Dutch teachers and in English for non-Dutch teachers ($n = 9$).

2.2.1. Demographics

The demographic questions addressed Gender, Age (years), Teaching Experience (years), Level of Education, Teaching Domain, and CT-experience. Answer options for *Level of Education*³ were: bachelor/master program at a university of applied sciences ($n = 54$), bachelor/master program at an academic university ($n = 182$), PhD ($n = 26$), and something else namely___ (one teacher reported vocational education as highest level of education). Answer options for *Teaching Domain* were (multiple answers possible): (1) technology; (2) ICT; (3) art & design; (4) economics & management; (5) welfare; (6) education; (7) health; and (8) law. We merged these into three broader domain-categories which paralleled the sections of the university of applied sciences: technology (category 1–3⁴ $n = 96$), economics (category 4; $n = 102$), and society (category 5–8; $n = 65$). Some teachers taught in multiple domains (8%): teachers in both technology and economics or society were assigned to the technological domain and teachers in both economics and society to the economical domain.

For *CT-experience*, teachers answered the question “Do you already have experience with CT? (multiple answers possible)”. Answer options were: (1) I took a CT course/workshop at this University of Applied Sciences; (2) I took a CT course/workshop somewhere else, namely___; (3) I developed a CT course for students; (4) I taught a CT course for students; (5) I am a member of the CT community of this University of Applied sciences; (6) I read a book about CT, namely___; and (7) something else, namely___. We assigned one point for each reported activity. Two raters coded all open answers, which were rewarded with one point if the reported activity addressed CT explicitly and did not belong to one of the already listed categories. Absolute agreement between the two coders on the total CT-experience score was 0.997 (two-way random effects intraclass-correlation coefficient; see Shrout & Fleiss, 1979). The two raters discussed the few inconsistent cases to reach consensus. The computed variable was a sum score of the number of reported activities, one extra point was added to this sum score if one of the reported activities involved – what we considered – deeper processing of CT-skills (i.e., teaching or developing a CT course, completing a bachelor/master in

philosophy, or conducting scientific research in the field CT). The possible score for CT-experience ranged from 0 to 8, with higher scores representing a higher level of CT-experience.

2.2.2. CRT performance

We measured CRT performance with a seven-item CRT developed by Toplak et al. (2014), that we translated into Dutch. Toplak et al. (2014) extended the original three-item CRT by Frederick (2005) to increase the reliability of the test and because the original items may have become familiar to participants. As explained (see section 1.1), the CRT is a short math test designed to measure the tendency to override an intuitive response-alternative that is incorrect and engage in further reflection to arrive at the correct response. An example item (Toplak et al., 2014) is: “Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? (intuitive answer: 30; correct answer: 29)”. Six items had a free-response format and one item had a multiple-choice format with three answer options. Correct answers were rewarded with one point, incorrect answers with zero points. CRT performance was computed as the sum score on all seven items and could therefore range from 0 to 7. Cronbach's alpha in the original study was 0.72 and in our sample it was 0.66.

2.2.2.1. CRT confidence and calibration. Participants rated the confidence in their response to each CRT-item by answering the question “How certain are you that your response is correct?” on a four-point rating scale ranging from (1) very uncertain; (2) somewhat uncertain; (3) somewhat certain; to (4) completely certain. CRT confidence was the average of the confidence ratings for all seven CRT-items and could therefore range from 1 to 4. To compute a calibration index, we identified the frequencies of the false negatives (i.e., very uncertain or somewhat uncertain judgment for a correct answer), true negatives (i.e., very uncertain or somewhat uncertain judgment for an incorrect answer), true positives (i.e., completely certain or somewhat certain judgment for a correct answer), and false positives (i.e., completely certain or somewhat certain for an incorrect answer) for each CRT-item separately.

2.2.3. Teaching attitudes

In order to measure teachers' attitudes towards teaching CT more generally, we constructed a questionnaire (see Table 1) that addressed teachers' (1) perceived relevance of teaching CT and perceived competence in teaching CT. Participants rated their agreement to 19 statements on a six-point rating scale ranging from (1) strongly disagree to (6) strongly agree. They received the following instruction “Suppose we would define critical thinking as follows: ‘critical thinking means that one engages in reflective reasoning before deciding what to believe or what to do, and that one can explain what those beliefs or decisions are based on’. Then to what extent do you agree with the following statements?” Thirteen of the 19 statements that followed were items from Stedman and Adams (2012; translated for the Dutch version), who measured perceptions of CT-instruction. Because the ability to avoid biases in reasoning and decision-making is an important element of CT, we added five items that addressed this element of CT as an addition to other CT elements that the items in the questionnaire addressed (see Table 1). We used one item as a control-item to check whether teachers adopted the frame of reference (see Table 1). Teachers should generally agree with this item if they had read the provided definition carefully, which was the case ($M = 4.8$, $SD = 1.0$). Nine of the 18 items intended to measure Perceived Relevance; the other nine Perceived Competence.

Because the questionnaire was a combination of self-constructed and not frequently used items, we examined the two-factor structure with a confirmatory factor analysis (CFA) using

³ In the Dutch education system, higher education can be higher professional education offered by universities of applied sciences (Bachelor, Master), and academic education offered by academic universities (Bachelor, Master, PhD, with the PhD being an additional four-year trajectory after a Master degree).

⁴ At the university of the present study, Art & Design belongs to the technology domain because these subjects are embedded in a more technology/ICT context (e.g., programming).

Table 1
Standardized and unstandardized coefficients for two-factor CFA on the 18-item teaching attitudes questionnaire.

Observed variable	Latent construct	β	<i>B</i>	<i>SE</i>
CT is essential in making important decisions (new)	Perceived Relevance	0.38	1.00	
CT during educational activities discourages students from active learning (reverse)	Perceived Relevance	−0.43	−1.43	0.38
CT allows students to better understand the course content (y1)	Perceived Relevance	0.63	1.46	0.33
CT during educational activities encourages students to become independent thinkers	Perceived Relevance	0.58	1.35	0.28
I believe it is more important for students to trust their intuition, than to evaluate evidence (reverse; new)	Perceived Relevance	−.27	−0.89	0.26
Learning outcomes will not improve from CT during educational activities (reverse; y2)	Perceived Relevance	−.58	−1.33	0.30
I believe that it is my responsibility to promote CT in my courses	Perceived Relevance	.53	1.11	0.30
CT is a way of thinking that would help students enjoy the learning process	Perceived Relevance	.62	1.55	0.35
CT helps students to see the difference between intuition and a balanced argument (y3; new)	Perceived Relevance	.53	0.98	0.23
While teaching, I look for specific evidence of CT by students	Perceived Competence	.62	1.00	
If required, I could implement CT into my courses (y4)	Perceived Competence	.72	1.06	0.13
I think that students have barriers to CT, regardless of the strategies I use (reverse)	Perceived Competence	−.12	−0.24	0.17
In order for me to fully implement CT in my courses I would need additional support (reverse)	Perceived Competence	−.56	−1.23	0.19
I find it hard to explain to my students why they are drawing incorrect conclusions from given information (reverse; new)	Perceived Competence	−.37	−0.75	0.19
I have the skills necessary to promote students' CT in my courses (y5)	Perceived Competence	.60	1.03	0.16
Usually, it is hard to determine whether students engage in CT during my courses (reverse)	Perceived Competence	−.58	−1.22	0.18
I am aware when students give intuitive answers during my lessons (new)	Perceived Competence	.26	0.41	0.13
I find it hard to integrate CT in the content I am teaching (reverse; y6)	Perceived Competence	−.69	−1.48	0.22
Critical thinking should always include a reflective component (control item)				

Note. $N = 254$. CFA = confirmatory factor analysis. Boldfaced variables are the indicators in the final measurement model in displayed in Fig. 2. New = newly constructed items.

the 'lavaan' package in R (R Development Core Team, 2008; Rosseel, 2012). We used a robust maximum likelihood (MLR) estimation as the assumption of multivariate normality was violated. Table 1 shows the standardized and unstandardized factor loadings. The model did not fit the data well, Comparative fit index (CFI) = 0.85, Root mean square error of approximation (RMSEA) = 0.063, and Standardized root mean square residual (SRMR) = 0.074. Based on the factor loadings in this model and reconsideration of the items' interpretation, we explored other factor structures. We found that a two-factor model with three items as indicators of Perceived Relevance and three of Perceived Competence (boldfaced items in Table 1) fitted the data well, CFI = 0.98, RMSEA = 0.054, SRMR = 0.035. The rationale behind the new item selection was that we selected only those items that addressed teachers' relevance perception of teaching CT for students' learning specifically, and only those that addressed teachers' perceived competence in teaching CT in one's courses. The excluded items of Perceived Relevance described other relevance aspects (i.e., relevance of CT in general, for enjoyment of learning, for active learning, or as personal responsibility to teach). The excluded items of Perceived Competence referred to a specific aspect of teaching CT (i.e., explaining or recognizing incorrect conclusions) or actually described behavior or faith in students' ability instead of competence perception.

The thin lines in Fig. 2 depict the final measurement model graphically, including the standardized factor loading and the squared multiple correlations (SMC) in italics. The SMSs indicate the (lower bound) reliability of the items. For example, Perceived Relevance accounted for 38% of the variance in y1. Item-correlations with means and standard deviations are shown in Table 2.

2.2.4. Thinking dispositions

Teachers' rational thinking dispositions were measured with two questionnaires: the 41-item Actively Open-minded Thinking scale (AOT; Stanovich & West, 2007) and the 18-item (short form) of the Need For Cognition scale (NFC; Cacioppo, Petty, & Feng Kao, 1984). Again we used translations for the Dutch version of the survey (derived from Heijltjes et al., 2014). Participants rated their agreement to the 59 statements in total on a six-point rating scale ranging from (1) strongly disagree to (6) strongly agree. Scores on the items were averaged for NFC and AOT separately (after reverse

scoring items that were formulated negatively) and could therefore range from 1 to 6. Higher scores on the AOT represent a stronger tendency towards open-minded thinking. An example item is "A person should always consider new possibilities." Cronbach's alpha was .82. Higher scores on the NFC represent a stronger tendency to engage in and enjoy thinking. An example item is "The notion of thinking abstractly is appealing to me." Cronbach's alpha was .85.

2.3. Analysis

To test our hypotheses, we computed a structural equation model (SEM) using the *lavaan* package in R (R Development Core Team, 2008; Rosseel, 2012). SEM is a combination of factor analysis (measurement model) and multiple regression (structural model). The measurement model examines relationship between the latent variables and their measures (i.e., answers to items). The structural model tests the interrelations among latent and observable variables. We included the measurement model of Perceived Relevance and Perceived Competence. We lacked power, however, to include the measurement models of all latent variables in our model; therefore, we included the latent variables that we measured with existing instruments (i.e., AOT, NFC, and CRT) as observed variables (mean centered). We used MLR estimation for our hypothesized model and a bootstrap estimation approach with 5000 samples to test the indirect effects in our explorative mediation analyses. Lastly, within our final sample of 263 teachers, 24 cases contained missing values: nine participants dropped out before starting the teaching attitudes questionnaire that followed after the CRT and, additionally, fifteen participants before starting the final questionnaire on thinking dispositions. Little's Missing Completely at Random (MCAR) test indicated that MCAR could be inferred, $\chi^2 = 32.91$, $df = 22$, $p = .063$. Therefore, we handled missingness using Full Information Maximum likelihood (FIML).

3. Results

To check whether the required assumptions for SEM were met, we checked for normality, outliers, and linearity. First, CRT performance and CRT confidence were negatively skewed; therefore we reflected both variables (i.e., subtracting each score from the largest score plus 1) and subsequently applied square root transformations (Tabachnick & Fidell, 2014). To enhance interpretation, we re-

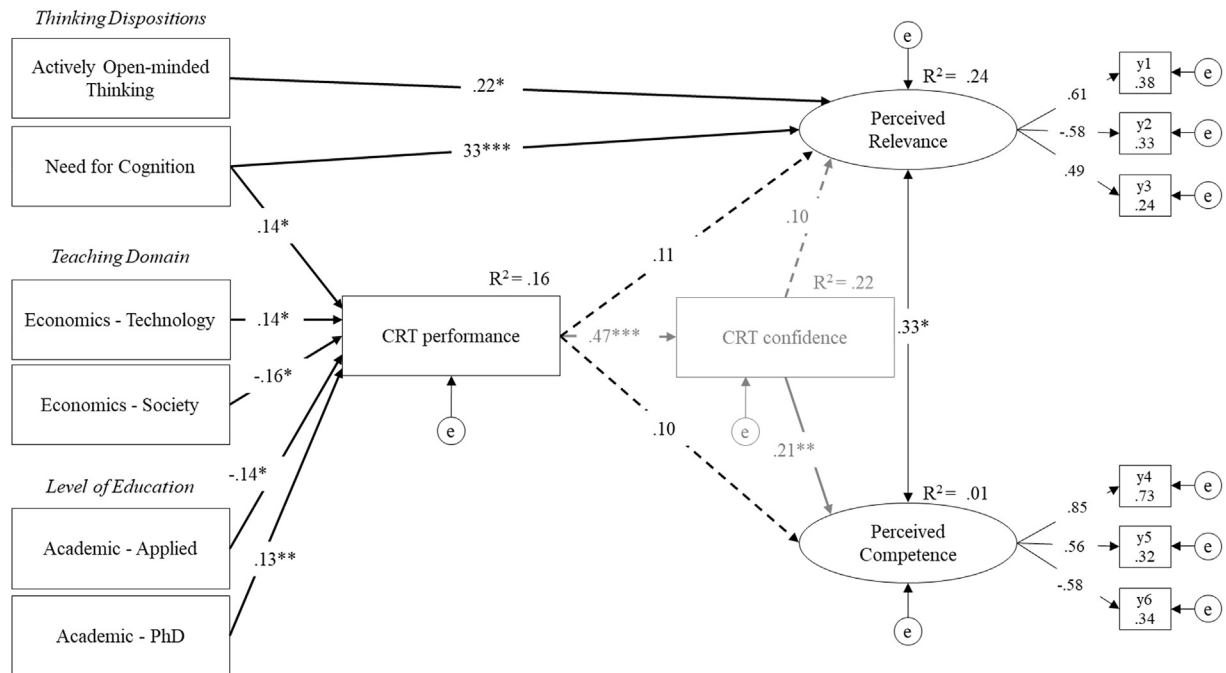


Fig. 2. Results of the structural equation model of thinking dispositions, Teaching Domain (dummies), Level of Education (dummies), CRT performance, CRT confidence, Perceived Relevance and Perceived Competence. Thin lines and the bolded represent the measurement component and structural component, respectively. Rectangles and circles represent observed and latent variables, respectively. Values in y1 to y6 are squared multiple correlations, indicating the reliability of each measure. Path coefficients are standardized regression weights. Dashed lines indicate nonsignificant paths. CRT = Cognitive Reflection Test. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 2
Means (M), standard deviations (SD), and correlations between items measuring perceived relevance and perceived competence.

	M (SD)	y1	y2	y3	y4	y5
y1	4.89 (0.82)					
y2	2.06 (0.82)	-.42***				
y3	5.17 (0.66)	.35***	-.32***			
y4	4.76 (0.78)	.36***	-.20**	.12		
y5	4.30 (0.91)	.25***	-.11	.16*	.49***	
y6	2.95 (1.13)	-.14*	.16*	-.001	-.51***	-.34***

Note. Range: 1–6. * $p < .05$. ** $p < .01$ *** $p < .001$.

reflected the variables after transformation. Second, we identified a total of eight univariate outliers; two on CRT confidence, two on NFC, and four on AOT, that we winsorized to fit the distribution (i.e., the difference between the two next highest or lowest values was added or subtracted to the next highest or lowest value with standardized value < 3.29 or > -3.29 ; Tabachnick & Fidell, 2014). We ran the analyses both with winsorized outliers and without outliers, yielding highly similar results (dissimilar results are reported). Third, in contrast to our hypotheses, there was no linear relationship between AOT and CRT; therefore this path was excluded from our structural model.

3.1. Descriptives

Table 3 displays Spearman correlations between all study variables and the means with standard deviations. The teachers obtained, on average, relatively high scores on CRT performance, CRT confidence, Perceived Relevance, Perceived Competence, AOT, and NFC. Regarding calibration, Table 4 shows that 73.5% of all CRT performance judgments were accurate, indicating that teachers were quite well calibrated. Interestingly, accurate judgments

mostly followed after a correct CRT performance (i.e. 1218 out of the 1354 accurate judgements concerned correctly performed CRT items), whereas the inaccurate judgments mostly followed after an incorrect CRT performance (i.e. 411 out of the 487 inaccurate judgements concerned in incorrectly performed).

The correlation analyses (Table 3) showed that CRT performance and CRT confidence were positively related ($r = 0.44$, $p < .001$), indicating that teachers with a better CRT performance were on average more confident about their performance. Those with a higher confidence in their CRT performance were, however, not necessarily better performers: despite higher CRT confidence ($r = 0.31$, $p < .001$), older teachers did not perform significantly better on the CRT than younger teachers (note, though, that more years of teaching experience was significantly positively related to CRT performance and confidence). Furthermore, teachers' need for cognition (NFC) was positively related to both their CRT performance ($r = 0.18$, $p = .005$) and CRT confidence ($r = 0.22$, $p = .001$), and also to their perceived relevance of teaching CT ($r = 0.34$, $p < .001$) of and perceived competence in teaching CT ($r = 0.15$, $p = .024$), and to amount of previous experience with CT-activities (CT-experience: $r = 0.15$, $p = .024$). Teachers' disposition towards actively open-minded thinking (AOT), was only significantly related to their perceived relevance of teaching CT ($r = 0.29$, $p < .001$) and CT-experience ($r = 0.17$, $p = .008$).

3.2. Structural equation model

Fig. 2 graphically displays the results of our SEM model including the thinking dispositions AOT and NFC, Teaching Domain (dummies), Level of Education (dummies), CRT performance, Perceived Relevance, and Perceived Competence. Our model fitted the data well, CFI = 0.94, RMSEA = 0.041, SRMR = 0.050. The individual pathways are described below.

Table 3
Means (M), standard deviations (SD), and spearman correlations between study variables.

	M (SD)	1	2	3	4	5	6	7	8	9
1. Gender (females)	41.44%									
2. Age (years)	46.34 (10.70)	-.27***								
3. Teaching Experience (years)	9.75 (8.73)	-.08	.57***							
4. CT-Experience (range 0–8)	0.71 (1.24)	.03	.05	.08						
5. CRT performance (range 0–7)	4.92 (1.78)	-.17**	.10	.17**	.05					
6. CRT confidence (range: 1–4)	3.46 (0.56)	-.27***	.31***	.24***	-.08	.44***				
7. Perceived Relevance (range: 1–6)	5.00 (0.57)	-.01	.04	-.05	.15*	.10	.15*			
8. Perceived Competence (range: 1–6)	4.37 (0.74)	-.06	.07	.09	.13*	.08	.17**	.25***		
9. AOT (range: 1–6)	4.68 (0.35)	-.09	.06	-.07	.17**	.05	.06	.29***	.13	
10. NFC (range: 1–6)	4.62 (0.54)	-.09	.06	-.07	.15*	.18**	.22**	.34***	.15*	.35***

Note. Means and standard deviations are computed from the untransformed variables, correlation analyses included the transformed variables. Gender: male = 0, female = 1. CRT = Cognitive Reflection Test, NFC = Need for Cognition, AOT = Actively Open-minded Thinking. Correlation analyses with excluded outliers instead of winsorized outliers did not yield any different results with regard to the direction or the significance of the correlations.

* $p < .05$. ** $p < .01$ *** < 0.001 .

Table 4
Calibration index for the cognitive reflection test (CRT).

	Confidence judgment		Performance				
			Correct		Incorrect		Total
			n	%	n	%	
Very uncertain	24	1.3	48	2.6	72	3.9	
	False negatives		True negatives				
Somewhat uncertain	52	2.8	88	4.8	140	7.6	
Somewhat certain	310	16.8	186	10.1	496	26.9	
	True positives		False positives				
Completely certain	908	49.3	225	12.2	1133	61.5	
Total	1294	70.2	547	29.7	1841	99.9	

Note. 263 teachers \times 7 CRT-items = 1841 performance judgements. Total percentage is not 100 because of rounding.

3.2.1. CRT performance

While taking the other predictors of the model into account, NFC was positively and significantly related to CRT performance ($\beta = 0.14$, $SE = 0.07$, $p = .037$, 95% $CI = 0.009, 0.278$).⁵ This indicated that teachers with a stronger disposition towards effortful thinking, achieved (on average) higher CRT performance, which was in line with our hypothesis 1a. Furthermore, both dummy variables of Teaching Domain were significantly related to CRT performance (Economics vs. Technology: $\beta = 0.14$, $SE = 0.07$, $p = .027$, 95% $CI = 0.016, 0.267$; Economics vs. Society: $\beta = -0.16$, $SE = 0.08$, $p = .016$, 95% $CI = -0.289, -0.030$), such that, as expected (hypothesis 1b), teachers in the domain of technology achieved the highest average score ($M = 5.49$, $SD = 1.42$), followed by teachers in the domain of economics ($M = 4.84$, $SD = 1.89$) and teachers in the domain of society ($M = 4.20$, $SD = 1.81$). Finally, the dummy variables of Level of Education were significantly related to CRT as well (Bachelor/Master at an Academic university vs. University of applied sciences: $\beta = -0.14$, $SE = 0.07$, $p = .017$, 95% $CI = -0.255, -0.025$; Bachelor/Master at an Academic university vs. PhD: $\beta = 0.13$, $SE = 0.08$, $p = .004$, 95% $CI = 0.041, 0.221$), revealing that teachers

with a PhD scored, on average, the highest ($M = 6.08$, $SD = 0.109$), followed by teachers with an academic Bachelor or Master degree ($M = 4.91$, $SD = 1.80$) and teachers with an applied-university degree ($M = 4.39$, $SD = 1.72$), which was in line with our hypothesis 1c. R^2 indicated that the predictors in the model explained 16% of the variability in teachers' CRT performance, which is a medium overall effect (Cohen, 1988).

3.2.2. Perceived relevance

A positive covariance between Perceived Relevance and Perceived Competence ($\beta = 0.33$, $SE = 0.04$, $p = .019$, 95% $CI = 0.024, 0.273$) showed that both teaching attitudes were moderately interrelated. Furthermore, the significant regression coefficients for both NFC ($\beta = 0.33$, $SE = 0.08$, $p < .001$, 95% $CI = 0.103, 0.299$) and AOT ($\beta = 0.22$, $SE = 0.14$, $p = .018$, 95% $CI = 0.023, 0.240$)⁶ indicated that teachers with stronger dispositions towards effortful and actively open-minded thinking indeed perceived teaching CT as more relevant (hypothesis 2a). CRT performance was positively but not significantly related to Perceived Relevance ($\beta = 0.11$, $SE = 0.09$, $p = .249$, 95% $CI = -0.045, 0.172$). Thus, in contrast to our

⁵ This path became nonsignificant in the analyses where we excluded the outliers: $\beta = 0.13$, $SE = 0.07$, $p = .050$, 95% $CI = 0.000, 0.268$

⁶ This path became nonsignificant in the analyses where we excluded outliers: $\beta = 0.18$, $SE = 0.15$, $p = .053$, 95% $CI = -0.001, 0.215$.

hypothesis 2b, teachers with higher CRT performance did not perceive teaching CT as more relevant. R^2 indicated that the predictors in the model explained 24% of the variability in teachers' perceived relevance of teaching CT, which is a large overall effect (Cohen, 1988).

3.2.3. Perceived competence

Just as with Perceived Relevance, a positive but nonsignificant regression coefficient for CRT performance ($\beta = 0.10$, $SE = 0.09$, $p = .139$, 95% $CI = -0.028, 0.200$) suggested that teachers who scored higher on the CRT did not perceive themselves as more competent in teaching CT, which was incongruent with our hypothesis 3. Not surprisingly, the predictors in the model explained almost no variance in teachers' competence perception towards teaching CT ($R^2 = 0.01$).

3.2.4. Mediation analyses

As mentioned in the introduction, we also tested whether CRT performance mediated the relationship between thinking dispositions and perceived relevance of teaching CT (see section 1.4). Additionally, we explored whether teachers' confidence in their CRT performance mediated the hypothesized relationship of CRT performance with Perceived Relevance and Perceived Competence (section 1.4.1). We conducted all mediation analyses in one explorative model (see grey paths in Fig. 2) which yielded a good data fit as well, $CFI = 0.94$, $RMSEA = 0.039$, $SRMR = 0.049$. Furthermore, the direction and significance of all previously described relationships remained the same.

First, we found no indirect effect of NFC via CRT performance on Perceived Relevance, ($\beta = 0.01$, $SE = 0.01$, $p = .575$, 95% $CI = -0.012, 0.027$), suggesting that CRT performance did not mediate the relationship between teachers' disposition towards effortful thinking and their perceived relevance of teaching CT. Note, however, that the results of the first model already rejected our hypothesis that CRT performance was significantly related to Perceived Relevance. Second, although we found a positive direct effect of CRT performance on CRT confidence ($\beta = 0.47$, $SE = 0.01$, $p < .001$, 95% $CI = 0.353, 0.580$), we found no direct effect of CRT confidence on Perceived Relevance ($\beta = 0.10$, $SE = 0.55$, $p = .262$, 95% $CI = -0.035, 0.174$), nor an indirect effect of CRT performance via CRT confidence on Perceived Relevance ($\beta = 0.05$, $SE = 0.04$, $p = .252$, 95% $CI = -0.017, 0.082$). Hence, neither teachers' CRT performance nor their confidence in that performance was related to how relevant they perceived teaching CT to be. In contrast, CRT confidence did have a direct effect on Perceived Competence ($\beta = 0.21$, $SE = 0.23$, $p = .004$, 95% $CI = 0.067, 0.307$) and CRT performance had an indirect effect via CRT confidence on Perceived Competence ($\beta = 0.10$, $SE = 0.05$, $p = .007$, 95% $CI = 0.030, 0.149$), indicating that – despite a non-significant overall effect of CRT performance on Perceived Competence – teachers who demonstrated better CRT performance were more confident about their CRT performance, and those teachers perceived themselves as more competent in teaching CT. Finally, the explained variance by the predictors in this explorative model remained more or less the same for CRT performance ($R^2 = 0.16$) and Perceived Relevance ($R^2 = 0.25$) as compared to our hypothesized model. For Perceived Competence, however it increased to 5% which is a small to medium effect (Cohen, 1988).

4. Discussion

Research on CT highlights the crucial role of the teacher, yet research on teachers' CT and attitudes towards teaching it is scarce. This study was the first to investigate what teacher characteristics are associated with teachers' CRT performance, which assesses an

important aspect of CT, and their attitudes towards teaching CT more generally. Our findings can inform future research on how to better equip higher education teachers for teaching CT.

4.1. CRT performance

As we hypothesized (Hypothesis 1a), teachers with a stronger disposition towards effortful thinking (NFC) indeed performed better on the CRT. In contrast, the disposition to engage in active open-minded thinking (AOT) was not related to CRT performance. Furthermore, it should be noted that the relationship between NFC and CRT was uncertain given the wide confidence interval for the regression coefficient and because its significance depended on including the (winsorized) outliers. Hence, we found no strong support for the hypothesized relationship between teachers' thinking dispositions and their CT-skills. This was surprising because previous studies in (mainly) student populations consistently showed that both NFC and AOT positively correlated with the CRT (Baron, Scott, Fincher, & Emlen Metz, 2015; Campitelli & Gerrans, 2014; Frederick, 2005; Pennycook et al., 2015; Szasz, Szollosi, Palfi, & Aczel, 2017; Thomson & Oppenheimer, 2016; Toplak et al., 2011, 2014).

We see multiple possible explanations for these divergent findings. One explanation may be that we did not measure our latent constructs sufficiently. As we lacked power to test a full measurement model, we did not include NFC, AOT, and CRT as latent variables in our model (as we did with perceived relevance and perceived competence). Hence, the model did not test the data fit of these measures and did not take their measurement errors into account. This may be especially problematic for the AOT because – despite its frequent use and in contrast to the CRT and NFC – the factor structure is somewhat unclear. Svedholm-Häkkinen and Lindeman (2018) recently showed that AOT was not a unidimensional construct, which may be problematic for the interpretation of the sum scores. Future research should focus on further validating the AOT. Nevertheless, measurement problems do not provide an explanation of why other studies consistently found a relation between thinking dispositions and CRT performance as these studies also used sum scores instead of measurement models.

Another potential explanation could be that this relationship does not apply to higher education teachers. However, a more likely explanation seems that our study sample was not representative. As participation was voluntary and the survey length relatively long, the teachers who finished the entire survey were probably very conscientious and/or already quite enthusiastic about CT. Hence, it is likely that we systematically over-sampled for teachers who had a stronger tendency to engage in and enjoy thinking and thus scored high on NFC. This self-selection bias could also have affected other variables in our study, since NFC has been shown to be a reliable predictor of performance on a wide range of CT-tasks and other thinking dispositions (Stanovich, 2011; Stanovich et al., 2016). Indeed, the teachers in our sample performed particularly well on the CRT and had relatively strong rational thinking dispositions: they scored more than thrice as high ($M = 4.9$) on the CRT compared to a sample of Canadian university students ($M = 1.5$ cf. Toplak et al., 2014) and had stronger rational thinking dispositions (AOT: $M = 4.7$ NFC: $M = 4.6$) than students of their own university of applied sciences reported in another study (AOT $M = 4.0$; NFC $M = 3.9$ cf. Heijltjes et al., 2015). Moreover, the negatively skewed distributed CRT performance suggested that a considerable proportion of the teachers performed at ceiling, which may have caused a restricted range of values that reduced the correlations or made them more dependent on outliers in the sample.

Returning to our hypotheses, we expected that teaching in a

technological domain would predict better CRT performance (Hypothesis 1b) than teaching in an economical or societal domain respectively, and this was indeed the case. Teaching domain was considered to reflect individual differences in available mindware, that is, the declarative knowledge and skills needed for correct reasoning. Our findings are in line with results of previous studies on mindware and CT (again mainly conducted with student populations) showing that numeracy skills, which play a more important role and are taught more explicitly in study programs in the technological domains, predicted performance on the CRT and other types of heuristics-and-biases tasks (Campitelli & Gerrans, 2014; Frederick, 2005; Klaczynski, 2014; Liberali, Reyna, Furlan, Stein, & Pardo, 2012; Szasz et al., 2017).

Level of Education was also related to CT-skills: academic teachers with a doctorate degree (PhD) achieved the highest CRT average, followed by academic teachers without a doctorate, and non-academic teachers (University of Applied Sciences), respectively. As level of education is typically associated with cognitive ability, this corresponds with previous findings that cognitive ability predicts performance on the CRT and other heuristics-and-biases tasks (Frederick, 2005; Klaczynski, 2014; West et al., 2008).

In addition, we explored how well teachers were calibrated, that is, how accurately they could judge their own CRT performance. The calibration index showed that most teachers were, overall, highly accurate in judging their CRT performance. This is perhaps not surprising, because teachers in our study also performed well on the CRT, and it were the correct performances in particular that were judged accurately. That they were somewhat overconfident, is shown by the finding that inaccurate judgments mainly pertained to incorrect performances (rather than erroneously discarding a correct answer), indicating that teachers did not detect their thinking error when they made one. Our findings seem to confirm Hattie's (2013) suggestion (concerning students) that individuals know what they know, but are less able to judge what they do not know.

4.2. Teaching attitudes

In line with our hypothesis 2a, we found a relationship between teachers' thinking dispositions (both NFC and AOT) and perceived relevance. Hence, teachers' personal thinking dispositions indeed played a role in how relevant they perceived teaching CT to be. This finding is in line with the proposition by Van Aalderen-Smeets and Walma van der Molen (2015) within the domain of science teaching, that personal attitudes (on science in general) positively influence professional attitudes (towards teaching it). In this light, thinking dispositions could also be seen as an expression of personal attitude on CT in general. However, just as with NFC and CRT, the relationship between AOT and perceived relevance was uncertain given the wide confidence interval for the regression coefficient and because its significance depended on including the (winsorized) outliers. Here, we also propose that self-selection bias or the measurement quality of the AOT-scale may explain the ambiguous relationship.

Rather surprisingly and in contrast to our hypotheses, teachers' CRT performance was not related to perceived relevance of (hypothesis 2b) or perceived competence in (hypothesis 3) teaching CT. In the literature, however, performance attainments are viewed as one of the principal sources of one's task value and competence judgments (Bandura, 1982; Eccles & Wigfield, 2002). A possible explanation for the lack of predictive value here may be that the CRT is not a good indicator of performance attainment in teaching CT. With the CRT we assessed a specific aspect of CT (i.e., rational thinking within a mathematical context) and related it to teachers' perceived relevance of and perceived competence in teaching CT

more in general. Given the positive correlation between previous CT-experience and perceived competence, it is possible that positive experiences in following courses on CT or in teaching CT would be better predictors of perceived relevance of and perceived competence in teaching than a CRT performance. Therefore, it might be interesting for future research to investigate indicators of performance attainment in teaching CT.

Another possible explanation for the nonsignificant relationship between CRT performance and perceived competence comes from our explorative mediation analyses, where we did find an indirect effect of CRT performance, via CRT confidence, on perceived competence. Hence, only those teachers with higher CRT scores and, subsequently, a higher confidence about that performance perceived themselves as more competent in teaching CT. This may imply that better CT-skills are related to higher perceived competence in teaching it, but only if one recognizes that one possesses the skill. This is in line with the expectancy value literature, which states that one's own interpretations of previous achievements are the antecedents of perceptions of competence (Eccles & Wigfield, 2002). The additional direct effect of CRT confidence on perceived competence (independent of actual performance) suggests that teachers' competence perception may be affected by personal traits as well. Finally, note that the relatively small amount of explained variance in teachers' perceived competence also suggests that other characteristics not considered in our model (e.g., personality traits or positive experience with teaching CT), may be more important.

4.3. Limitations and implications for future research

The results of this study have to be seen in light of some limitations. First, given the self-selection bias that likely occurred, the results of this study should be interpreted with caution. We expect the means for CRT performance, thinking dispositions, and teaching attitudes in the general teacher population to be lower than observed in the current study. Furthermore, although we suspect that the self-selection bias probably reduced the size of the correlations in our study and that, consequently, the effect sizes for the studied relationships may be even larger in a more representative teacher sample (see section 4.1), we cannot know whether this is true. This is something that further research should point out. Nevertheless, most of our findings on the studied relationships were in line with previous findings in student populations, which seems to suggest they are meaningful.

A second potential limitation concerns our measurement instruments. We only focused on one (albeit important) aspect of CT (rational thinking) and regarded the CRT-score as a proxy of teachers' ability to avoid bias in reasoning and decision-making. Even though the CRT has been studied extensively and has been shown to be a reliable predictor of a person's ability to make unbiased judgments and rational decisions in a wide variety of contexts ((Pennycook et al., 2015; Primi et al., 2016; Toplak et al., 2011, 2014), it remains an open question whether our findings would also apply to other CT-tasks. Also note that the reliability of the CRT in our sample was somewhat lower than reported in the original study (in our study $\alpha = 0.66$ versus $\alpha = 0.72$ in Toplak et al., 2014). We suspect that this low reliability can be explained by low variance, due to the very high CRT performance in our study. Furthermore, all of our instruments were Dutch translations of English questionnaires, and although we do not expect cultural differences to impact the findings, we cannot fully rule this out, because a direct comparison has not been made. For the AOT and NFC, we used existing translations that had been used in previous research (e.g., Heijltjes et al., 2014) in which Dutch students achieved similar averaged sum scores on the translated questionnaires compared to

a US student sample (Stanovich & West, 2007) and, as in the US sample, their sum scores correlated positively with performance on heuristics-and-biases tasks. A final limitation is, as with any correlational design, that this study does not allow us to draw conclusions about causality or the directions of the studied relationships.

Despite these limitations, our findings have some interesting implications for future research on how to better equip higher education teachers for teaching CT. First, despite some ambiguous results, our findings underline that thinking dispositions are important to take into account when designing and investigating the effectiveness of interventions aimed towards improving teachers' CT-skills, attitudes, and ability to teach CT. An interesting open question for future research to address would be to what extent dispositions are malleable. For instance, the finding that CT-experience correlated positively with both dispositions (AOT and NFC) could imply that increasing experience with CT (e.g., through workshops, lectures, et cetera) would positively affect dispositions towards CT. On the other hand, teachers with stronger dispositions may have sought out more opportunities for engaging with CT. If dispositions would be malleable, however, an important question would be whether changing a teacher's disposition towards effortful and actively open-minded thinking would lead to better performance on CT-tasks and a more positive teaching attitude, which together would ultimately affect teaching behavior and quality.

Second, our findings regarding teaching domain and level of education seem to endorse the important role of mindware and cognitive ability in CT-skills. Hence, when training teachers, it is important to take their mindware into account and to address potential knowledge gaps. In our study, numeracy was important. However, for other types of CT-tasks, different kinds of mindware may be required (e.g., rules of logic, probabilities, et cetera) and may need to be explicitly trained. Yet having mindware available does not guarantee its application, and dispositions may again be important here. Klaczynski (2014) found that only at high levels of thinking dispositions, numeracy mindware was predictive of performance on the CRT and four other types of heuristics-and-biases tasks. Put differently, it seems that possessing mindware (or possessing great cognitive capacity) is only beneficial when you are also favorably disposed towards thinking critically. Thus, when training teachers it can indeed be helpful to take mindware into account, but it is important to keep in mind that possessing relevant mindware is not equal to being less prone to biases in reasoning and decision-making. Finally, our finding that better CRT performance was not directly related to a more positive attitude towards teaching CT, may suggest that training teachers' CT-skills do not automatically reinforce their teaching attitudes. Therefore, future research should experimentally investigate the effects of CT-training on both teachers' CT-skills and their attitudes towards teaching it.

In conclusion, by identifying variables that play a role in higher education teachers' CT-skills and teaching attitudes, the results of this study provide a first step towards future research on how to equip teachers for the important task of teaching CT. Future research should establish whether interventions targeting these variables would help to improve teachers' CT-skills, their teaching of CT and, ultimately, students' CT-skills.

Acknowledgement

The authors would like to thank Steven Raaijmakers for his assistance with the data analysis. This work was supported by the Netherlands Organisation for Scientific Research (project number 409-15-203).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tate.2019.05.008>.

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85, 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, 1102–1134. <https://doi.org/10.3102/0034654308326084>.
- Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, 24, 1–3. <https://doi.org/10.1016/j.learninstruc.2012.10.003>.
- Arum, R., & Roksa, J. (2011). Limited learning on college campuses. *Society*, 48, 203–207. <https://doi.org/10.1007/s12115-011-9417-8>.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122–147. <https://doi.org/10.1037/0003-066X.37.2.122>.
- Baron, J., Scott, S., Fincher, K., & Erlen Metz, S. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, 4, 265–284. <https://doi.org/10.1016/j.jarmac.2014.09.003>.
- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61, 1139–1160. <https://doi.org/10.1177/0018726708094863>.
- Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of Need for Cognition. *Journal of Personality Assessment*, 48, 306–307. https://doi.org/10.1207/s15327752jpa4803_13.
- Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, 42, 434–447. <https://doi.org/10.3758/s13421-013-0367-9>.
- Choy, S. C., & Cheah, P. K. (2009). Teacher perceptions of critical thinking among students and its influence on higher education. *International Journal of Teaching and Learning in Higher Education*, 20, 198–206.
- Cohen, J. (1988). In reprint (Ed.), *Statistical power analysis for the behavioral sciences* (p. 2). New York, NY: Psychology Press.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>.
- Evans, J. S. B. T., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A*, 54, 935–958. <https://doi.org/10.1080/02724980042000417>.
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19, 25–42. <https://doi.org/10.1257/089533005775196732>.
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73, 349–360. <https://doi.org/10.1093/poq/nfp031>.
- Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction*, 24, 62–66. <https://doi.org/10.1016/j.learninstruc.2012.05.009>.
- Heijltjes, A., Van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning and Instruction*, 29, 31–42. <https://doi.org/10.1016/j.learninstruc.2013.07.003>.
- Heijltjes, A., Van Gog, T., Leppink, J., & Paas, F. (2015). Unraveling the effects of critical thinking instructions, practice, and self-explanation on students' reasoning performance. *Instructional Science*, 43, 487–506. <https://doi.org/10.1007/s11251-015-9347-8>.
- Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21, 1161–1166. <https://doi.org/10.1177/01461672952111004>.
- Jones, A., & Moreland, J. (2003). Considering pedagogical content knowledge in the context of research on teaching: An example from technology. *Waikato Journal of Education*, 9. <https://doi.org/10.15663/wjve.v9i0.387>.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Los Angeles, California: Cambridge University Press.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Lane.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251. <https://doi.org/10.1037/h0034747>.
- Klaczynski, P. A. (2014). Heuristics and biases: Interactions among numeracy, ability, and reflectiveness predict normative responding. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00665>.
- Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 59–76. <https://doi.org/10.1016/j.edurev.2014.06.001>.

- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, 25, 361–381. <https://doi.org/10.1002/bdm.752>.
- Lunn, P. D. (2013). The role of decision-making biases in Ireland's banking crisis. *Irish Political Studies*, 28, 563–590. <https://doi.org/10.1080/07907184.2012.742068>.
- Marcus, B., Bosnjak, M., Lindner, S., Pilitschenko, S., & Schütz, A. (2007). Compensating for low topic interest and long surveys: A field experiment on nonresponse in web surveys. *Social Science Computer Review*, 25, 372–383. <https://doi.org/10.1177/0894439307297606>.
- National Research Council. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. In J. W. Pellegrino, & M. L. Hilton (Eds.), *Board on testing and assessment and board on science education, division of behavioral and social sciences and education*. Washington, DC: The National Academies Press.
- Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How robust are the findings of Academically Adrift? Change. *The Magazine of Higher Learning*, 43, 20–24. <https://doi.org/10.1080/00091383.2011.568898>.
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2015). Is the Cognitive Reflection Test a measure of both reflection and intuition? *Behavior Research Methods*, 48, 341–348. <https://doi.org/10.3758/s13428-015-0576-1>.
- Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research*, 42, 237–249. <https://doi.org/10.1080/001318800440579>.
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, 29, 453–469. <https://doi.org/10.1002/bdm.1883>.
- R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>.
- Ritchhart, R., & Perkins, D. N. (2005). Learning to think: The challenges of teaching thinking. In K. J. Holyoak, & R. G. Morrison (Eds.), *The cambridge handbook of thinking and reasoning*. Los Angeles, California: Cambridge University Press.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48. <https://doi.org/10.18637/jss.v048.i02>.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645–665. <https://doi.org/10.1017/S0140525X00003435>.
- Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford University Press.
- Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, 13, 225–247. <https://doi.org/10.1080/13546780600780796>.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking*. Cambridge, Massachusetts London, England: MIT Press.
- Stedman, N. L. P., & Adams, B. L. (2012). Identifying faculty's knowledge of critical thinking concepts and perceptions of critical thinking instruction in higher education. *Nacta Journal*, 56, 9–14.
- Svedholm-Häkkinen, A. M., & Lindeman, M. (2018). Actively open-minded thinking: Development of a shortened scale and disentangling attitudes towards knowledge and people. *Thinking & Reasoning*, 24(1), 21–40. <https://doi.org/10.1080/13546783.2017.1378723>.
- Szaszi, B., Szollosi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited: Exploring the ways individuals solve the test. *Thinking & Reasoning*, 23, 207–234. <https://doi.org/10.1080/13546783.2017.1292954>.
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (Pearson new international edition 6th ed.). Harlow: Pearson.
- Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and defense attorney's fallacy. *Law and Human Behavior*, 11, 167–187. <https://doi.org/10.2307/1393631>.
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11, 99–113.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39, 1275–1289. <https://doi.org/10.3758/s13421-011-0104-1>.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20, 147–168. <https://doi.org/10.1080/13546783.2013.844729>.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2017). Real-world correlates of performance on heuristics and biases tasks in a community sample: Heuristics and biases tasks and outcomes. *Journal of Behavioral Decision Making*, 30, 541–554. <https://doi.org/10.1002/bdm.1973>.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>.
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59, S251–S278. <https://doi.org/10.1086/296365>.
- Van Aalderen-Smeets, S. I., & Walma van der Molen, J. H. (2013). Measuring primary teachers' attitudes toward teaching science: Development of the Dimensions of Attitude toward Science (DAS) instrument. *International Journal of Science Education*, 35, 577–600. <https://doi.org/10.1080/09500693.2012.755576>.
- Van Aalderen-Smeets, S. I., & Walma van der Molen, J. H. (2015). Improving primary teachers' attitudes toward science by attitude-focused professional development. *Journal of Research in Science Teaching*, 52, 710–734. <https://doi.org/10.1002/tea.21218>.
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281. <https://doi.org/10.1080/14640746808400161>.
- 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, 930–941. <https://doi.org/10.1037/a0012842>.