Is Critical Thinking a Mediator Variable of Student Performance in School?

Christel Walter
University of Cooperative Education in Saxonia

Paul Walter
University of Bremen

The study explores the influences of critical thinking and interests on students' performance at school. The tested students attended German grammar schools ("Gymnasien").

Separate regression analyses showed the expected moderate positive influences of critical thinking and interests on school performance. But analyzed simultaneously, applying a Structural Equation Model, only a direct effect of critical thinking on school performance was observed.

Furthermore, critical thinking seems to be a moderator variable, mediating an indirect effect of interest on school performance. An additional analysis of the data showed that the influence of critical thinking could exclusively be observed in the subsample of students who had a family background without a migration history. In the subsample with migration history critical thinking and interests did not have an effect on school performance. Since the students with migration history did not differ in school performance from their fellow students without migration history, the result gives rise to the assumption that those students in German grammar schools may have chosen other ways of motivation and learning style to school performance.

Further research was done to clarify why critical thinking turns out to be a mediator of school performance and if this function is observable regarding other variables and different age groups. Beyond that we suggest reconsidering some theoretical and empirical issues, especially reviewing the relationship between critical thinking skills and dispositions.

In the USA, to boost critical thinking (CT) was declared a national priority in higher education some decades before (Halpern, 1998). A few years ago, the Association of American Colleges and Universities has published transdisciplinary guidelines to the valid assessment of learning in undergraduate education, among which one can find a CT rubric (AACU, 2010). Regarding the current research literature, several studies showed positive effects of trained CT on school performance (e. g. Heinrich, Habron, Johnson, & Goralnik, 2015). The educational importance seems to be beyond dispute, despite open questions. In our study some of the questions are explored hoping to stimulate further research. The topics are: After delineating known definitions and approaches of CT, conjectures about the relation of CT skills and CT dispositions made in several approaches; the universality vs. cultural sensitivity of the CT concept; the development of CT; and empirical findings about the impact of CT on school performance.

To begin with a common and widely shared understanding of CT and its objectives one can cite Ennis's definition (2011): CT "is reasonable and reflective thinking focused on deciding what to believe or do". To place CT into the framework of higher order cognitive skills, Jonassen (2000) considered CT as part of complex thinking processes, besides "content thinking" (acquisition of accepted knowledge) and creative thinking (generating new knowledge). In the model, CT has to reorganize knowledge using the skills of analyzing, evaluating, and connecting information. Facione (2013) distinguished six "core" CT skills: interpretation, analysis, inference, evaluation, explanation, and self-regulation. These overlapping skills are the result of a Delphi method based on a panel of experts and also name the subtests of the California Critical Thinking Skills Test developed by the author and his co-workers. Similar skills (supposing a larger or a smaller number of skills)

can be found in other CT approaches or tests. It is a characteristic of broad concepts like CT that there are different approaches containing similarities and differences, we cannot discuss in the actual paper.

Concerning the assumption of separable CT skills and dispositions Facione (2013) added several CT dispositions to the above-mentioned skills: He characterized a "strong critical thinker" as inquisitive, judicious, truth seeking, confident in reasoning, open-minded, analytical, and systematic. Besides 26 different CT skills, Paul and his co-workers distinguished the following CT traits or virtues (e.g. Paul, Binker, Martin, Vetrano, & Kreklau, 2010, p. 58): Intellectual humility, intellectual courage, empathy, integrity, perseverance, sense of justice, and faith in reason. This classification apparently includes elements of theories of personality and values.

Regarded from an empirical standpoint, both classifications are difficult to evaluate. Facione and his co-workers have tried to measure their supposed seven dispositions in a questionnaire (the California Critical Thinking Dispositions Inventory). But his taxonomy was regarded as over-inclusive (Black, 2008) and the test dimensions were empirically not completely confirmed (Kakai, 2003). It can be added that Facione reported weak or moderate correlations (.09 to .41) of CT skills and dispositions in different samples (Facione, 2000).

The attempts to measure CT skills and dispositions separately are not fully conclusive. For example, why did Facione define "self-regulation" as a skill, "systematicity" as a CT disposition? The same doubts pertain for "analysis" and "analytic". A reason for the partitioning may be the different assessment methods of dispositions and skills. Skills are usually assessed by psychometric tests, dispositions by questionnaires. However, CT tests do not exclusively measure skills. For instance, some items of the known Watson-Glaser Critical Thinking Appraisal (WGCTA) (Watson & Glaser,

1980) demand the ability to make correct inferences on practical issues. Such competencies in deductive logic are also required in intelligence tests. But several other items of the WGCTA presuppose a positive attitude to accept the most obvious or the most probable of the offered alternative answers. To some extent, the results of the test reflect subjective preferences, which conform to the beliefs of the test authors.

Nevertheless, there are at least two plausible reasons for the distinction of CT skills and dispositions. The first reason is based on the known fact that intellectual virtues and attitudes are a necessary condition for the (adequate) use of CT skills: If you have to decide in a controversial issue – e.g. ecologically vs. economically compatible solutions of mass unemployment – you can try to get the best solution by analyzing different reasons with perseverance; or you may opt for the proposal of an authority without further consideration, even if you have the necessary cognitive tools for reasoning. One can see in this case that motivational and ethical dispositions are important for the use or non-use of CT.

The theoretical cause of the postulation of CT skills *and* dispositions lies in history. Both Facione and Paul located the origin of CT in the Socratic philosophy, which has been further developed in the subsequent Western thinking. The fundamental idea of the Socratic criticism is the use of *and* the trust in rationality. To solve complex problems is not per se critical in the Socratic sense of the concept. To be critical in this sense, a responsible citizen has to use intellectual skills, which are supported by the strong obligation to question common assumptions and to come to a decision on the basis of sound arguments and ethical values. In other words, skills and dispositions originally are two inseparable sides of criticism.

A *methodological* explanation of the disputable relation of CT skills and dispositions was not made to the present, as far as we know. Following the classical distinction between typical and maximal behavior by Cronbach (1949), one could characterize CT dispositions as typical behavior, CT skills as maximal behavior. Typical behavior is usually measured by self-reports of abilities or preferences, while maximal behavior is assessed by psychometric tests. By the way, the research on intelligence shows that the demands of the situation are crucial, whether typical or maximal performance has a greater impact (Wilhelm & Engle 2005). But unlike the field of intelligence research, the available measures of CT cannot absolutely separate skills and dispositions. The methodological difficulty to distinguish maximal and typical CT points to a conceptual characteristic. Both have such a strong connection that they may be separated theoretically as skills and dispositions, but their measurement as isolated variables causes difficulties. CT tests may assess the *potential* maximum of CT, but cannot predict CT of a person working on a specific task.

The origin in the Socratic and Western philosophy gives rise to the further question, whether CT is a universal concept of thinking. Some authors differentiated between a Socratic learning style and a Confucian learning tradition (Perez & Shin, 2016; Tweed & Lehman, 2002). To what amount the Confucian tradition is responsible for Asian students' successful learning and high scores in large scale assessments like PISA (Program for International Student Assessment) (OECD, 2014), remains an open question. But the Confucian tradition may be only one alternative to Socratic criticism. Other common sense learning styles or implicit adaptation techniques may exist, which are not yet discussed in philosophy or science. On the other hand, we can find publications about CT in the educational literature all over the world, also in Eastern Asia; but the supposed

universality may limit the spectrum of the concept and be a cause that there the majority of publications refers to CT *skills*.

The next point to mention is the small empirical knowledge about the ontogenetic development of CT. King and Kitchener (2004) presented a developmental model of reflective judgment, an integral but just a specific component of CT. Their model described the developmental progression by seven stages, ranging from pre-reflective to reflective thinking. Reflective judgment about ill-structured issues (e.g. safety of chemical additives to foods) was characterized by the evaluation of evidence or opinion from different perspectives or across similar contexts. The reflective judgment model is based on a large number of interviews of adolescents and adults in longitudinal and cross-sectional designs, going back till the 1970s.

On the contrary, several components of CT are taken into account by Elder and Paul (2010) in their proposal of six developmental stages; at the beginning of the development stands the "unreflective thinker"; the "accomplished thinker" is located on the highest stage six. The authors described what CT skills and traits are predominating and lacking on each of these stages. The authors asserted that their model is "based on the nearly twenty years of research" (Elder & Paul, 2010, "Critical Thinking Development", para. 2) and instruction experiences, but also conceded that their approach reflected an "intellectual" not a "psychological standpoint". This statement could indicate an only loose coupling of research results to the development model.

Furthermore, the existing developmental models of CT seem to be top down constructed: Lower stages are defined as deficient modes of elaborated CT or as lack of CT components. Such a modeling does not support an educational goal to better CT. In instructional settings the knowledge would be helpful which thinking processes can be

regarded as *preforms* of CT; if so, teachers could and should foster such preforms.

Despite the open questions, CT skills are treated as important factors in educational contexts. Studies consistently report positive correlations between CT skills and achievement scores, such as grades, GPA or school graduations (Facione, 2013; Sourisseaux, Felsing, Müller, Stübig, Schmücker, & Heyde, 2007; Watson & Glaser, 1980). In many studies, CT assessments serve as dependent variables to evaluate the outcome of CT trainings (e.g. Adair & Jaeger, 2016). Looking at sites of the internet, CT training is recommended in an immense amount simply presupposing its usefulness.

The findings in regard to the CT dispositions are less obvious: The correlations between CT dispositions and achievement scores are significant, but rather moderate (e.g. Karagöl & Bekmezci, 2015 reported a typical correlation of $r = .17$ in a sample of Turkish teacher candidates) and lower than the corresponding correlations of CT skills. The different correlations can be caused by real differences in the influence of CT skills and dispositions or by the circumstance that typical and maximal CT behavior is not clearly weighted in the applied assessment methods. Furthermore it has to be clarified, whether the different magnitudes of the reported correlations reflect characteristics of the tested samples (with regard to age, variance of CT scores, etc.).

A matter of concern is the question, how the relationship of CT to achievement variables will change, if other variables are simultaneously measured. For instance, if variables like interest, intelligence, etc. have been considered in a study, it is possible that the effects of CT may decrease or be partialled out. In other words, the variance of the factor CT would then be explained by other better defined variables such as interest, intelligence or achievement motivation. CT could lose its educational relevance, or its role reduces to a

general but dispensable term. Studies which deal with such a precarious problem are rare. An exception is the study of Aksu and Koruklu (2015), who investigated effects on students' math success in Turkish high schools. Scores of a CT disposition questionnaire were correlated with math success in a moderate magnitude (r = .18). The math grades were also positively correlated with logical thinking and attitude towards mathematics. When modeling the attitude towards mathematics as a mediator variable, the direct effect of CT on math success lost statistical significance. Only an indirect effect of CT dispositions mediated by the attitude towards mathematics attained statistical significance. Focusing CT dispositions may have added to this result.

This attenuation effect in mind concerning CT we wanted to know whether we have to face a similar risk in our context. Such a negative result might reduce the importance of our general background objective to foster CT in schools. Therefore we examined the effect of CT on school success, when an additional predictor was simultaneously observed. The variable *interest* was selected as a known motivational and volitional factor to school success and as a potential objective of instruction. In addition, we tested whether CT has the same importance in students of families coming from different countries and cultures. We had to limit our study to an age group of students, who have probably not developed full CT. Though our study should contribute to the question, whether we can assess preforms of "strong" CT and what are their characteristics. This is a prerequisite task, if the fostering of CT in childhood and adolescence is regarded as an educationally meaningful objective.

Method

Participants

113 students of four German grammar schools (Gymnasien) participated in the study. All students belonged to the 7th

grade (mean age: 12.93 years). 63 students were female, 50 students male; the sex difference reflected the higher proportion of female students in German higher education (Quenzel & Hurrelmann 2010).

We deliberately picked grammar schools out, where we could expect a large number of students with a migration background. As a consequence 59 students in our sample were foreign born (n = 7) or had at least one foreign born parent. However all students in the sample attended a German school from the beginning. The sample comprised students with parents coming from 14 states (24 from eastern European countries, 21 from Turkey, 4 from Eastern Asia states, 10 from other European and other states), a typical distribution pattern of the migrant population in Germany. The selected schools were located in districts of a big German city with a major rate of lower-income families.

Measures

Critical thinking

There are few assessment methods of CT in Germany. They are not applicable in the examined age group, not to mention conceptual shortcomings. Therefore we designed a technique tailored to the sample, a first step to a more sophisticated method. Different components and also preforms of CT, thinking skills and value-oriented, ethical reasoning were to be assessed. Probably more typical than maximal CT was demanded in three written scenes administered as tasks related to the school subjects math, German studies, and biology. The scenes should stimulate the students to critical statements. The "math-item" starts with a teacher question "How is it possible, to get the sum of six throwing two dice?" The response of the fictitious student was "six and zero". The students had to comment on the irritating response of the fictitious student in writing. The "German studies item" was taken from a short story of the writer Gianni Rodari,

who tells a surrealistic version of the Pied Piper of Hamelin substituting the former rats for cars. The students had to comment on a woman's illogical statement in the story: "There are so many cars in the city that they don't have space either to drive or to stop. They can only be pushed away!" The third item stemmed from a lesson in biology. Two sentences were presented concerning the behavior of lions: "The lions are waiting, observing, and selecting the easiest prey", and "The lions unexpectedly attack and hunt the animals which are to slow and cannot escape". A fictitious student judged the two sentences as contradictory. The students had to explain if and why the fictitious student was right or wrong.

The 339 student answers were rated by the two researchers and signed by codes representing CT components. Before coding, the researchers carefully studied definitions and illustrations of CT components in articles of above-mentioned authors: Ennis (2011), Facione (2013), Halpern (1998), Paul et al. (2010). That way a conceptual validity should be achieved for the selected and rated CT components. The components explanation, interpretation, open-mindedness, intellectual courage, truth-seeking, maturity of judgment, and intellectual clarity could be distinguished in the student answers. Manifestations of inferences or analyses of relations were not observed in any student answer.

On each CT component, all answers were rated on a scale from 0 (component is lacking) to 2 (high manifestation of the component). The factual coding procedure began, when the coding training had reached about 80% agreement on each of the CT component. Applying the principle of consensual coding, controversial codes were discussed to find a common solution. 13 items (of 3x7 items) formed a sufficient reliable scale (Cronbach's $\alpha = .82$), which was used in the statistical analyses. This means that CT could be detected in the students' answers in varying degrees.

Academic interest

The students' academic interests were assessed by a subscale of the "Potsdam Motivation Inventory" (PMI), which is used in educational and psychological research. We used one dimension of the multi-dimensional construct interest. The PMI subscale (interest in math) was applied in a similar sample (5 to 9 graders) by Rheinberg and Wendland (2002). They reported a reliability of $\alpha = .81$. The correlation to the later math grade averaged $r = .27$.

We constructed our own scales for interest in German studies and in biology (eight items each) analogous to the PMI math subscale (example: "It would be hard for me if we had no biology lessons at school"). The reliability of the scales was sufficient (math: $\alpha = .87$; biology: $\alpha = .84$; German studies: $\alpha = .83$).

Grades

The students' last school grades in math, biology, German studies served as criteria variables (in Germany, grades are expressed as integers from 1 = very good to 6 = inadequate). It was our aim to explore their dependency on CT and/or academic interest. It is to mention that the division of dependent and independent variables is not strict since the students' interest and CT scores may partly depend on previous school achievement scores.

Results

At first, we looked at the relationships between the three variable groups. In bivariate analyses, we computed the correlations among the predictors (i.e. between interest and the CT scores) and did separate simple regression analyses of each predictor and the respective grade-variable. We used CT

as a compound variable which consisted of the score on the reliable scale of 13 items. Figure 1 contains the results of these analyses.

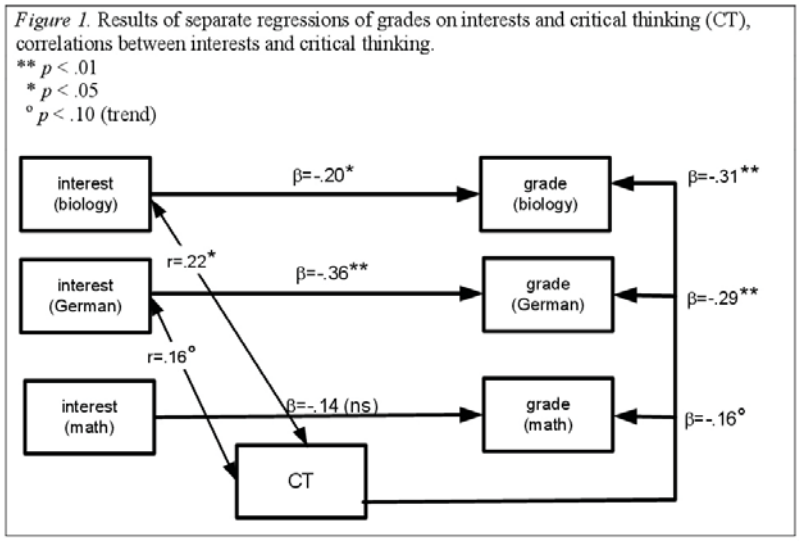


Figure 1. Results of separate regressions of grades on interests and critical thinking (CT), correlations between interests and critical thinking.
** $p < .01$
* $p < .05$
° $p < .10$ (trend)

Critical thinking, interests, and grades

Interests as well as CT predicted the students' grades, but not for math (minus signs of the standardized regression coefficients β were caused by the fact that lower grades signified better performance). The missing significance of the regressions on math grades may be due to sample characteristics (restricted sample size; less female students' interest in math).

Separate multiple regressions (interest and CT as predictors of the three dependent variables) showed similar but some additional results (Table 1).

Table 1: Prediction of grades in separate multiple regression analyses

Variable	*B*	95% CI
	Grade: biology[a]	
Constant	3.48**	[2.81, 4.16]
Interest in biology	-0.17	[-0.39, 0.50]
Critical thinking	-0.07**	[-0.12, -0.03]
R^2	0.12	
F	6.90**	
	Grade: math[b]	
Constant	3.59**	[2.83, 4.35]
Interest in math	-0.16	[-0.37, 0.56]
Critical thinking	-0.04[o]	[-0.10, 0.07]
R^2	0.04	
F	2.54[o]	
	Grade: German[a]	
Constant	4.23**	[3.52, 4.93]
Interest in German studies	-0.37**	[-0.57, -0.17]
Critical thinking	-0.06**	[-0.10, -0.02]
R^2	0.18	
F	12.28**	

Note. **CI = confidence interval.**
[a]***N*** **= 112.** [b]***N*** **= 113.**
****p*** **< .01.** *__p__ **< .05.** [o]***p*** **< .10.**

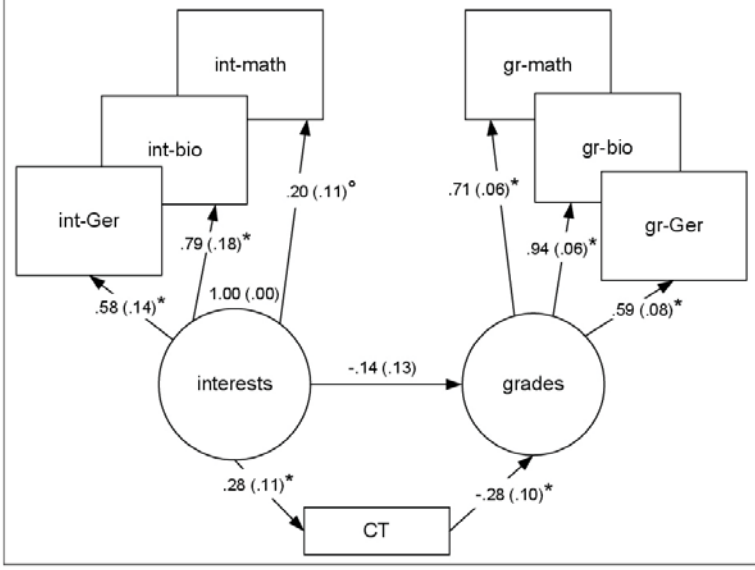
Different from the results of simple regression, the interest in biology did not predict biology grades in the multiple regression. The loss of statistical significance was caused by the correlation among the two predictors (s. Figure 1), besides restriction of sample size. Table 1 also shows the differences of the explained variances of the three dependent variables by the predictors: The predicted variance of the grades in German studies was remarkable high (R^2 = 0.18); the corresponding value of biology grades was moderate (R^2 = 0.12), of math grades only marginal (R^2 = 0.04).

As the observed different regression results on biology grades suggest, CT may be regarded as a moderator for interest. To examine this conjecture the influence of interests on grades had to be divided in direct and indirect effects.

The analysis was done with the *SEM*-Algorithm of Mplus (Muthén &Muthén, 2012), due to the limited sample size in a heuristic manner. The result is displayed in Figure 2. The data of grades and interests were used to estimate latent variables *interests* and *grades*. The standardized parameter estimates on both latent factors served as weights in a later analysis. The parameter estimates of the latent factors are acceptable except for the interest in math.

Indirect effect of interests

Figure 2. SEM-Model of school performance, interests, and critical thinking, using standardized model statistics.
$N = 110$
* $p < .05$
° $p < .10$
indirect effect (interests – CT – grades) = -.08 ($p = .06$)



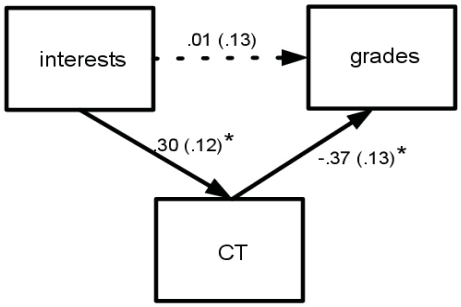
While CT had a significant influence on grades, interests did not show a direct effect on grades. But there was a trend towards an indirect effect of interests on grades. The indirect influence just missed statistical significance level (Figure 2). Interests seemed to be related to good school performance, only if students showed CT.

The role of cultural background

We further examined the possible dependency of the CT scores on the students' cultural background. The CT scores of students with and without family migration background showed a significant difference. Students without migration background scored higher than their fellow students with migration background: M_1 = 7.07 (SD_1 4.53), M_2 = 5.24 (SD_2 2.98) (t[111] = 2.52 $p < .05$; eta^2 = .06.). However, both groups did not differ in the scores for (pooled) interests and grades. In the case of grades as of CT, the standard deviations were smaller in the student group with family migration background.

An additional examination of the data consisted in computing path analyses. The analyses demonstrated that the pattern of results for the *whole* group was completely determined by the subgroup of students *without* a migration story. Figure 3 presents the results for this subgroup (using standardized estimates). CT predicted school grades to a substantial degree in the subgroup without migration history. In contrast, interests were associated with school performance, only if mediated by CT. The indirect effect of interests almost reached significance level. The restricted sample size was responsible for the marginal missing of this level (valid N = 52 of this subsample): In the path analysis for the total sample, the indirect effect of interest for grades reached statistical significance.

Figure 3. Path analysis of school performance, interest, and critical thinking in students without migration history, using standardized statistics.
N = 52
* p < .05
indirect effect (interests – CT – grades) = -.11 (*p* < .07)



Discussion

CT is fostering school performance in a moderate degree. The result resembles that reported in the literature. Similar results are obtained for interests in separate regression analyses. The effect of CT does not disappear, when academic interests are simultaneously measured. Contrary to the result of Aksu and Koruklu (2015), CT becomes a mediator of students' interests. Only if students develop higher levels of CT, interests have an impact on school performance. This result may depend on the objectives of the German school system, especially of grammar schools. These schools put emphasis on self-regulated learning and on the development of students' autonomy. CT is favorable to these school demands and a precondition to activate the influence of academic interests on school performance. The observed small, only indirect influence of interests on school performance may also be caused by the known weakening of interests in school subjects, when students are at the age of secondary education (Krapp, 2001). A lack of interest in school topics may be compensated by CT, which can function as a stabilizing factor for school performance. Whether particular components of CT are responsible for the mediation effect cannot be said on the basis of the results. But the major demand of typical CT and the lack of some higher order skills in the students' answers suggest that motivational and attitudinal components of CT played an important role.

There are some caveats to add in interpreting our result which, at the same time, can be read as recommendations for future research on CT. First, the impact of CT on school performance and as a mediator of interests was observed only in the subgroup of students *without* migration history. CT is lower in the subgroup *with* migration background and their interests in school topic are not activated by CT. Despite comparable school performance,

the students with migration background in our sample seem to be not optimally prepared for the objectives of the German school system that enhances the principle of self-directed learning and that is therefore demanding CT on the student side. Perhaps, an assimilation pressure in the families with migration background induces a high extrinsic orientation in school success or achievement motivation in their children so that their school performance does not need to decrease. To clarify possible different effects in different ethnic groups, other variables must be examined. In the literature of intercultural education we would find potential "candidates" (e.g. individualism-collectivism, achievement motives).

Second, the estimated effects of CT on school performance are small, its function as mediator is a bit arbitrary. The moderate effect size may be due to the assessed components and also due to shortcomings of the assessment technique. Both problems have to do with the discussed difficulty to separately assess CT dispositions and skills, typical and maximal thinking processes. Modern approaches of multitrait-multimethod matrix-analysis (i.e. models of confirmatory factor analysis; Geiser, Eid, Nussbeck, Lischetzke, & Cole, 2010) are applicable to handle this difficulty. In a first step, applying different CT measures, such analyses could test what measures constitute CT factors and what measures constitute method factors. In this way one would gain some insights to relevant CT components and to the methods that fit best to assess these components. The next step of confirmatory factor analysis could be to analyze relevant factors of CT (found in the first step) within the network of convergent and discriminant variables. Only such a complex research strategy can give a profound empirical answer to the question whether CT can be seen as a concept sui generis. Only then can the mediator role of CT be thoroughly investigated.

A characteristic of our study was the small age span in the sample. In this age group, we could assess predominantly preforms of CT. On the one hand the students showed some CT in different degrees; on the other hand we did not find any manifestations of inferences or analytical thinking when the students worked on the presented tasks. More sophisticated tasks as in scientific or political contexts would demand qualitative higher CT than the assessed thinking processes of the sample students. Only cross-sectional designs with different age group or longitudinal designs do allow a check if a prediction of CT on school performance is similar at different ages. Particularly in the case of the sample students who had not yet shown completely elaborated forms of CT, such an assumption would be risky. Furthermore, developmental research would be desirable to improve the instruction of CT in schools. It would help teachers to know their students' preforms and forms of CT, they could foster.

Our recommendations laid emphasis on further empirical research of CT. But there is also a lot of theoretical work to do. For instance, the relation of skills and dispositions or typical and maximal CT is far from a sufficient theoretical reconstruction. CT – in our opinion – should not be conceived as a mere instrumental tool for predetermined goals but also as a value-oriented thinking style. Perhaps theoretical work will reveal something like a conceptual basis of CT as general mediator to handle and solve complex intellectual and social problems.

References

Association of American Colleges and Universities. (2010). Critical thinking value rubric. Retrieved from https://www.aacu.org/value/rubrics/critical-thinking

Adair, D., & Jaeger, M. (2016). Incorporating critical thinking into an engineering undergraduate learning environment. *International Journal of Higher Education, 5*, 23-39. doi:10.5430/ijhe.v5np23.

Aksu, G., & Koruklu, N. (2015). Determination the effects of vocational high school students' logical and critical thinking skills on mathematics success. *Eurasian Journal of Educational Research, 59*, 181-206. Retrieved from http://dx.doi.org/10.14689/ejer.2015.59.11

Black, B. (2008). Critical thinking – A definition and taxonomy for Cambridge assessment: Supporting validity arguments about critical thinking assessments administered by Cambridge assessment. Paper presented at 34th International Association of Educational Assessment Annual Conference, Cambridge, September 9.

Cronbach, L. J. (1949). *Essentials of psychological testing.* New York, NY: Harper & Row.

Elder, L., & Paul, R. (2010). Critical thinking development: A stage theory. Retrieved from http://www.criticalthinking.org/pages/critical-thinking-development-a-stage-theory/483

Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. Retrieved from http://www//faculty.education.illinois.edu/rhennis/documents/TheNatureofCriticalThinking_51711_000.pdf

Facione, P.A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic, 20*, 61-84.

Facione, P.A. (2013*). Critical thinking: What it is and why it counts*. Milbrae, CA: California Academic Press.

Geiser, C., Eid, M., Nussbeck, F.W., Lischetzke, T., & Cole, D.A. (2010). Multitrait-Multimethod-Analyse. In H. Holling & Schmitz, B. (Eds.), *Handbuch statistik, methoden und evaluation* (pp. 679-685). Göttingen: Hogrefe.

Halpern, D. (1998). Teaching critical thinking for transfer across domains. *American Psychologist, 53*, 449-455.

Heinrich, W.F., Habron, G.B., Johnson, H.L., & Goralnik, L. (2015). Critical thinking assessment across four sustainability-related experiential learning settings. *Journal of Experiential Education 38*, 373-393. doi: 10.1177/1053825915592890.

Jonassen, D.H. (2000). *Computers in the classrooms. Mindtools for critical thinking*. Englewood Cliffs, NJ: Prentice Hall.

Kakai, H. (2003). Re-examining the factor structure of the California Critical Thinking Disposition Inventory. *Perceptual and Motor Skills, 96*, 435-438.

Karagöl, I., & Bekmezci, S. (2015). Investigating academic achievements and critical thinking dispositions of teacher candidates. *Journal of Education and Training Studies 3*, 86-92. doi:10.11114/jets.v3i4.834.

King, P. M., & Kitchener, K. S. (2004). Reflective judgment. Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist 39*: 5-18.

Krapp, A. (2001). Interesse. In D.H. Rost (Ed.), *Handwörterbuch pädagogische psychologie* (pp. 286-294). Weinheim: Psychologie Verlags Union.

Muthén, L. K., & Muthén, B. O. (2012). Mplus, version 7.3. Los Angeles, CA: Muthén and Muthèn.

Organisation for Economic Co-operation and Development. (2014). PISA 2012 results in focus. Retrieved from

https://www.oecd.org/pisa/ keyfindings /pisa-2013-results-overview.pdf

Quenzel, G., & Hurrelmann, K. (2010). Geschlecht und Schulerfolg: Ein soziales Stratifikationsmuster kehrt sich um. *Kölner Zeitschrift für Soziologie, 62*, 61-91.

Paul, R., Binker, A.J.A., Martin, D., Vetrano, C., & Kreklau, H. (2010). *Critical thinking handbook: 6th – 9th Grades. A guide for remodeling lesson plans in language arts, social studies, and science.* Dillon Beach, CA: Foundation f. Critical Thinking Press.

Perez, A., & Shin, M. H. (2016). Study on learning styles and Confucian culture. *Indian Journal of Science and Technology, 9,* 1-5. doi: 10.17485/ijst/2016/ v9i26/97395.

Rheinberg, F., & Wendland, M. (2002). Veränderung der Lernmotivation in Mathematik: Eine Komponentenanalyse auf der Sekundarstufe I. *Zeitschrift f. Pädagogik, 45*. Beiheft, 308-320.

Sourisseaux, A., Felsing, T., Müller, C., Stübig, S., Schmücker, J., & Heyde, G. (2007). *Deutsche adaptation des critical thinking Appraisal von G. Watson und E. Glaser.* Frankfurt a. M., Germany: Harcourt.

Tweed, R. G., & Lehman, D. R. (2002). Learning considered within a cultural context. Confucian and Socratic approaches. *American Psychologist, 57*, 89-99.

Watson, G., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal.* San Antonio, TX: The Psychological Corporation.

Wilhelm, O., & Engle, R. W. (2005). Intelligence: A diva and a workhorse. In R. W. Engle, & Wilhelm, O. (Eds.), *Handbook of understanding and measuring intelligence* (pp. 1-9). Thousand Oaks, CA: Sage Publ.