

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Hyytinen, Heidi; Nissinen, Kari; Kleemola, Katri; Ursin, Jani; Toom, Auli

Title: How do self-regulation and effort in test-taking contribute to undergraduate students' critical thinking performance?

Year: 2024

Version: Published version

Copyright: © 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Hyytinen, H., Nissinen, K., Kleemola, K., Ursin, J., & Toom, A. (2024). How do self-regulation and effort in test-taking contribute to undergraduate students' critical thinking performance?. *Studies in Higher Education*, 49(1), 192-205. <https://doi.org/10.1080/03075079.2023.2227207>



How do self-regulation and effort in test-taking contribute to undergraduate students' critical thinking performance?

Heidi Hyytinen, Kari Nissinen, Katri Kleemola, Jani Ursin & Auli Toom

To cite this article: Heidi Hyytinen, Kari Nissinen, Katri Kleemola, Jani Ursin & Auli Toom (2023): How do self-regulation and effort in test-taking contribute to undergraduate students' critical thinking performance?, *Studies in Higher Education*, DOI: [10.1080/03075079.2023.2227207](https://doi.org/10.1080/03075079.2023.2227207)

To link to this article: <https://doi.org/10.1080/03075079.2023.2227207>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 Jun 2023.



Submit your article to this journal [↗](#)



Article views: 226



View related articles [↗](#)



View Crossmark data [↗](#)

How do self-regulation and effort in test-taking contribute to undergraduate students' critical thinking performance?

Heidi Hyytinen ^a, Kari Nissinen ^b, Katri Kleemola ^a, Jani Ursin ^b and Auli Toom ^a

^aCentre for University Teaching and Learning, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland; ^bFinnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT

Critical thinking is a multifaceted construct involving a set of skills and affective dispositions together with self-regulation. The aim of this study was to explore how self-regulation and effort in test-taking contribute to undergraduate students' performance in critical thinking assessment. The data were collected in 18 higher education institutions in Finland. A total of 2402 undergraduate students at the initial and final stages of their bachelor degree programmes participated in the study. An open-ended performance task, namely the Collegiate Learning Assessment (CLA+) International, was assigned to assess students' critical thinking, and a self-report questionnaire was used to measure self-regulation and effort in test-taking. Information on test-taking time was also utilised in the analysis. The interrelations between the variables were analysed with correlations and structural equation models. The results indicate that self-regulation in test-taking has only indirect effects on critical thinking performance task scores, with effort and time as mediating variables. More precisely, planning contributed to critical thinking performance indirectly through test-taking time and effort, while monitoring had no significant relation to critical thinking performance. The findings did not differ between the initial-stage and final-stage students. The model explained a total of 36% of the variation in the critical thinking performance task scores for the initial-stage students and 27% for the final-stage students. The findings indicate that performance-based assessments should be carefully designed and implemented to better capture the multifaceted nature of critical thinking.

ARTICLE HISTORY

Received 28 October 2022
Accepted 14 June 2023

KEYWORDS

Critical thinking; self-regulation in test-taking; performance task; test-taking effort; test-taking time; performance-based assessment

1. Introduction

Critical thinking can be seen as comprising two dimensions, namely a set of skills and affective dispositions in using those skills to guide behaviour and to decide what to believe (Ennis 1996; Facione 1990; Halpern 2014). Although critical thinking is conceptualised as a multifaceted construct in the literature, contemporary empirical research on critical thinking tends to focus more on the specific skills instead of covering the complexity of this phenomenon (Bensley et al. 2016; Rear 2019). This is not surprising since the assessment of such a complex and multifaceted construct is not without its challenges.

In recent years, assessment of critical thinking has increasingly moved towards authentic performance-based assessment (Braun et al. 2020; Shavelson et al. 2019; Tremblay, Lalancette, and

CONTACT Heidi Hyytinen  heidi.m.hyytinen@helsinki.fi  Centre for University Teaching and Learning, Faculty of Educational Sciences, University of Helsinki, Helsinki 00014, Finland

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Roseveare 2012). Performance-based assessment refers to assessment tasks that resemble the real-world situations in which critical thinking is required (McClelland 1973; Shavelson et al. 2019). In order to capture the complex nature of critical thinking, such assessment should require not only certain skills, but also considerable effort, willingness and a tendency to use critical thinking skills (Braun 2019; Ercikan and Oliveri 2016; Kane, Crooks, and Cohen 2005). Additionally, to succeed in performance-based assessment, students should also be capable of self-regulating; in other words, they should be able to plan and monitor their thoughts and behaviours regarding the demands of the task (Bensley et al. 2016; Hyytinen et al. 2021). However, surprisingly little attention has been paid to the role of self-regulation and test-taking effort in critical thinking performance in the context of performance-based assessment (Hyytinen et al. 2021; Ghanizadeh and Mizae 2012).

To address this gap, this study investigates how self-regulation and effort in test-taking contribute to undergraduate students' critical thinking performance in a performance-based assessment, namely the *Collegiate Learning Assessment (CLA+) International*. The findings contribute to developing performance-based critical thinking assessments further and to drawing more valid inferences about students' critical thinking performance.

1.1. Assessing critical thinking with performance-based assessments

Critical thinking is conceptualised as open-minded, purposeful and self-regulative thinking that combines a set of skills and affective dispositions to assess the reliability of information, to find alternative solutions, to recognise biases, to consider various perspectives and the range of possible consequences, as well as to reach a well-reasoned conclusion (Ennis 1996; Facione 1990; Halpern 2014; Shavelson et al. 2019). Affective dispositions, sometimes called 'critical spirit' or 'intellectual commitment', embody a multitude of dimensions that refer to individuals' willingness and tendency to use critical thinking skills (Ennis 1996; Halpern 2014; Holma 2016). The following dispositions are emphasised in the literature: open-mindedness, persistence, diligence, and effort (Bensley et al. 2016; Facione 1990; Halpern 2014). Dispositions are crucial in the assessment of critical thinking. It has been suggested that students who have strong affective dispositions are much more likely to use their critical thinking skills in various situations than those students who have mastered the skills but are not disposed to use them (Facione 1990). However, it is important to note that affective dispositions are strongly connected to the contextual elements of the assessment situation (Bensley et al. 2016). For instance, low-stakes critical thinking assessments that have no consequences may not strongly encourage students to put their best efforts into completing the assessment tasks (cf. Silm, Pedaste, and Täht 2020; Wise, Kuhfeld, and Soland 2019).

There is evidence showing that there is considerable variation in critical thinking among undergraduate students (Badcock, Pattison, and Harris 2010; Evens, Verburgh, and Elen 2013; Kleemola, Hyytinen, and Toom 2022a; Utriainen et al. 2017), and that the development of critical thinking skills is limited (Arum and Roksa 2011). Most studies have focused on critical thinking skills while dispositions have received less attention (Bensley et al. 2016; Rear 2019). Critical thinking skills are traditionally assessed using self-reports that actually reveal little about the actual skills of a student (Zlatkin-Troitschanskaia, Shavelson, and Kuhn 2015). Performance-based assessments, such as open-ended performance tasks and multiple-choice tasks could be a better option here and they have indeed become increasingly popular in the assessment of critical thinking. These task types focus on different skill sets, and it has been found that performance tasks activate more holistic use of skills compared to multiple-choice tasks (Hyytinen et al. 2021; Davey et al. 2015; Kleemola, Hyytinen, and Toom 2022a). Thus, open-ended tasks are considered appropriate in the assessment of complex constructs such as critical thinking. Performance tasks tend to simulate authentic problem-solving situations, which allow assessing, evaluating, synthesising, and interpreting relevant knowledge associated with a situation, as well as using that knowledge to solve a problem,

and communicating their response in writing (Davey et al. 2015; Braun et al. 2020; Shavelson et al. 2019). In line with this conceptualisation, we focus on three aspects of critical thinking: analysis and problem-solving (i.e. the ability to utilise, analyse, and evaluate the information provided and the ability to reach a conclusion), writing effectiveness (i.e. the ability to elaborate and to provide arguments that are well-constructed and logical), and writing mechanics (i.e. the ability to produce a well-structured text; see Kleemola, Hyytinen, and Toom 2022a).

Performance tasks have been criticised due to their emphasis on communicative skills, such as writing (Aloisi and Callaghan 2018). While communicative skills are often omitted from the definition of critical thinking, the ability to convey one's thinking to others is vital, for instance through argumentation (Kuhn 2019; Halpern 2014). Consequently, writing or oral communication can be seen functionally as a part of critical thinking (see Kleemola, Hyytinen, and Toom 2022a; 2022b). In the assessment of critical thinking, particularly in performance tasks, writing skills inevitably influence the outcome, as students are required to provide their responses in the form of a written report. Therefore, the possible interference of writing skills needs to be acknowledged when assessment findings are interpreted (Kleemola, Hyytinen, and Toom 2022a), but this should not entail excluding communicative skills from critical thinking assessments altogether.

1.2. Self-regulation and its associations with critical thinking

Self-regulation refers to the intentional management of one's own learning that allows students to plan, set goals, control effort, and monitor their thoughts, emotions, and behaviours according to the demands of the learning situation or task (Boekaerts and Cascallar 2006; Zimmerman 2002; Schunk and Greene 2018). Previous literature has suggested that self-regulation includes three intertwined phases: planning, monitoring, and evaluation (Usher and Schunk 2018; Zimmerman 2002). Planning refers to a student's awareness of the task, setting goals and analysing the task, and identifying the approaches and strategies that will be needed to accomplish the task (Toering et al. 2012). Monitoring refers to the active phase of observing performance within a given context, and modifying and adapting strategies according to the demands of the task (Usher and Schunk 2018). During and after the performance phase, students need to evaluate and review their effort, behaviours and strategies to decide whether their performance meets the goals (Zimmerman 2002). In particular, self-regulation is seen as essential when it comes to demanding learning events and tasks, in which students need to actively plan and monitor their activities and progress (Hoyle and Dent 2018; Hyytinen et al. 2021; Saariaho et al. 2019; Toering et al. 2012). Research has shown that there is considerable variation in undergraduate students' self-regulation skills (Räisänen, Postareff, and Lindblom-Ylänne 2016).

Although theorisation of critical thinking has established a link between critical thinking and self-regulation (Facione 1990; Halpern 2014), research investigating these two phenomena together is sparse (Bensley et al. 2016; Ghanizadeh and Mizae 2012; Hyytinen et al. 2021). It follows that the association between critical thinking and self-regulation is still not clear. On the one hand, self-regulation is sometimes considered a critical thinking skill (Facione 1990). On the other hand, self-regulation of learning is also viewed as a central guiding element of the complex process of critical thinking and use of various skills (Hyytinen et al. 2021; Ghanizadeh and Mizae 2012; Maksum, Widiana, and Marini 2021; Uzuntiryaki-Kondakci and Çapa-Aydin 2013). However, it is important to note that previous research on critical thinking has mostly focused on students' self-regulatory learning processes, but seldom on self-regulation applied in critical thinking assessments. Respectively, self-regulation research in higher education has focused more on learning processes and emotions than on investigating associations between self-regulation and critical thinking (e.g. Schunk and Greene 2018). Therefore, more research is needed in clarifying the relationship between critical thinking and self-regulation, how they are intertwined (Bensley et al. 2016), and how they contribute to each other. In particular, there is a need for research that examines self-regulation exercised in critical thinking assessment (Hyytinen et al. 2021).

1.3. Test-taking effort, time, and self-regulation as predictors of critical thinking test performance

The assessment of critical thinking requires the presence of the corresponding affective dispositions, such as effort (Bensley et al. 2016; Ennis 1996; Facione 1990). With respect to the nexus between test-taking effort and test performance, researchers have found significant interrelationships between these variables. For instance, it has been found that test-taking effort (i.e. how much effort students put into test-taking) has a significant and substantial impact on critical thinking test performance (Liu et al. 2016; Bensley et al. 2016). A similar positive association has been found between general test performance and test-taking effort among higher education students in low-stakes assessments (Silm, Must, and Täht 2019; Silm, Pedaste, and Täht 2020; Wise, Kuhfeld, and Soland 2019).

For the most part, test-taking effort is measured using self-report instruments. However, using actual test-taking time (i.e. response time) as an indicator of test-taking effort has increased. The effect of self-reported effort on test performance has been found to be less pronounced than using response time as a measure of test-taking effort (Wise, Kuhfeld, and Soland 2019). Nevertheless, it seems that these two measures of effort (i.e. self-reported and time-based) complement each other (Silm, Must, and Täht 2019). In this study, we use both test-taking time and self-reported effort as measures of test-taking effort.

Furthermore, as noted above, responding to performance-based critical thinking assessment involves self-regulation skills, such as abilities to plan, set goals, control effort, and monitor thoughts, emotions, and behaviours according to the demands of the assessment task (Hyytinen et al. 2021; Boekaerts and Cascallar 2006). Earlier research suggests that self-regulation of learning is a significant predictor of critical thinking among high school students, as assessed using self-reports (Gurcay and Ferah 2018). A few studies have also suggested that the association between monitoring learning processes and critical thinking is reciprocal, as measured by self-report surveys or multiple-choice questionnaires among higher education students (Bagheri and Ghanizadeh 2016; Ghanizadeh and Mizae 2012). It seems that self-regulation supports complex and multidimensional critical thinking (Bensley et al. 2016; Facione 1990; Ghanizadeh and Mizae 2012). Thus, together with test-taking effort, self-regulation should have an influence on students' test performance. While the association between critical thinking and self-regulation has been investigated to some degree, it is important to note that there is very little research on self-regulation in the context of performance-based assessment.

The present study provides new insights into this relatively little-explored aspect of critical thinking research by examining interconnections with higher education students' critical thinking performance, test-taking effort, and self-regulation applied in the test situation in the context of performance-based critical thinking assessment.

2. Research questions

In the study at hand, our aim is to investigate associations between critical thinking performance, self-regulation, and test-taking effort among undergraduate students in the CLA + International performance-based assessment (see Section 3.3 *Statistical analyses* and Figure 1 for further clarification of the structural equation model). We explore the associations at two stages of undergraduate-level studies, namely at the initial and final stages of their studies. More precisely, we pose the following research questions:

- (1) What is the relationship between self-regulation in test-taking and critical thinking performance task scores among initial- and final-stage undergraduate students?
- (2) What is the role of test-taking effort and time in self-regulation and critical thinking performance task scores among initial- and final-stage undergraduate students?

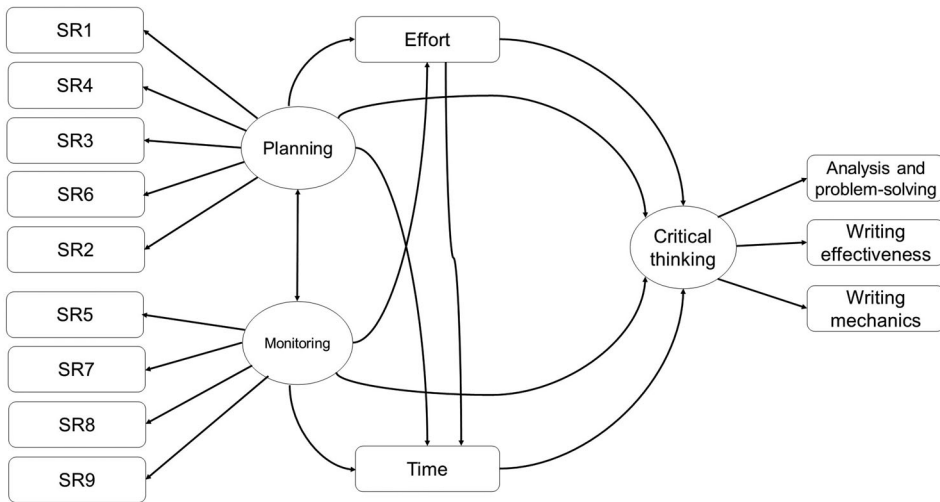


Figure 1. The structural equation model of the relationships between planning, monitoring, effort, time and critical thinking.

3. Methods and materials

3.1. Context, participants, and data collection

The Finnish higher education system consists of 38 institutions with around 300,000 students in total. The evaluation in Finnish higher education is 'enhancement-oriented', whereby the aim is to further elevate the quality of the programmes and institutions. Consequently, Finland has not adopted a culture of testing in developing education, and assessments in Finnish higher education are typically low-stakes as opposed to high-stakes. This applies to all levels of education (Ursin 2020).

The target group for this study consisted of undergraduate students at the initial (first-year) and final (third-year) stages of their undergraduate degree programmes at 18 higher education institutions. The data collection was conducted by cluster sampling, whereby the study programmes of selected higher education institutions served as clusters. It should be noted that participating in the study was voluntary for the institutions. Thus, strictly speaking, the data cannot be considered a genuine random sample from the Finnish student population. It is important to keep this in mind when performing statistical inference. However, to improve the national representativeness of the data, we stratified the student population of the 18 voluntary higher education institutions according to fields of study (following the international ISCED 2011 classification of broad fields of education) and sampled the participating study programmes within these strata, across the institutions. This ensured that all broad fields available in the Finnish higher education institutions were represented in the collected data. In the statistical analyses, the distribution of the broad fields in the data was corrected to match that of the Finnish student population by applying survey weights.

In small programmes, all initial- or final-stage students were selected, while in programmes with a large number of students, the participating students were drawn randomly. In all, 2402 undergraduate students participated in the study. Of these, 1538 (64%) were initial-stage students and 864 (36%) final-stage students. The participants consisted of 1178 (49%) males, 1158 (48%) females and 66 (2%) individuals who did not wish to state their gender. The participation rate was 25%. The participation of individual students was also voluntary, and informed consent was obtained from all participants. The data collection was part of a national project funded by the Ministry of Education and Culture.

3.2. Measures

We measured critical thinking with a performance task (henceforth PT) selected from the computer-assisted instrument CLA + International (Collegiate Learning Assessment; Zahner and Cioffi 2018). In the PT, students were asked to produce a written answer to an open-ended problem which dealt with differences in life expectancies in two cities, which mimics a complex real-world situation with the need for a well-founded conclusion or decision (Ursin and Hyytinen 2022). In order to successfully complete the PT, students needed to familiarise themselves with the reference materials available in the task and base their answers on these. Students had 60 min to complete the PT, which measured critical thinking skills in three categories: analysis and problem-solving (APS), writing effectiveness (WE), and writing mechanics (WM). Each student's answer was scored by two independent and trained scorers according to the CLA + scoring rubric (Kleemola, Hyytinen, and Toom 2022b). The score for each component ranged from 0 to 6 points, where 0 was given to responses which showed no effort to answer the question at all. Measured by correlation, agreement between the two scorers was acceptable (APS $r = 0.71$, WE $r = 0.71$, WM $r = 0.70$). Scoring consistency was also controlled by monitoring the given scores. If the two scores given to a response differed by more than two points, a third scorer was employed and the two closest scores were retained. It has been shown that the scores of the three skill components load strongly onto a latent variable measuring critical thinking (Kleemola, Hyytinen, and Toom 2022a).

After completing the test, the students filled in a questionnaire for background information purposes. Self-regulation in test-taking was measured with questions about planning and monitoring exercised when working with the PT. We employed nine items adapted from the Self-Regulation of Learning Self-Report Scale (SRL-SRS; Toering et al. 2012) to match the testing situation. The answers were given on a five-point scale (1 = fully disagree, 2 = disagree, 3 = neither agree or disagree, 4 = agree, 5 = fully agree). The items were expected to form two scales of self-regulation: planning (5 items) and monitoring (4 items). The items and the postulated scales are presented in Table 1, which also includes basic descriptive statistics of the items measuring planning and monitoring. The statistics are reported separately for initial- and final-stage students.

We evaluated the feasibility of the two scales of self-regulation in the current data, together with the factor of critical thinking, by conducting confirmatory factor analysis. An excellent fit was observed. The fit indices were 0.04 for the Root Mean Square Error of Approximation (RMSEA), 0.98 for the Comparative Fit Index (CFI), and 0.04 for the Standardised Root Mean Square Residual (SRMR; see Hu and Bentler 1999). The reliability estimates of the scales were 0.79 for planning, 0.79 for monitoring, and 0.86 for critical thinking.

The test-taking effort of students was measured with a question about how much effort the student had put into completing the PT. The effort was reported on a five-point scale (1 = no effort at all, 2 = a little effort, 3 = a moderate amount of effort, 4 = a lot of effort, 5 = my best

Table 1. Descriptive statistics of items of planning and monitoring in test-taking.

Scales	Items	Initial-stage students			Final-stage students		
		n	Mean	SD	n	Mean	SD
Planning	I developed a plan for solving the problem. (SR1)	1436	3.38	0.90	815	3.47	1.17
	I went through the steps of the plan I had made to solve the problem in my mind. (SR4)	1427	3.35	0.90	808	3.37	1.19
	I imagined the parts of the problem I still had to complete. (SR3)	1423	3.70	0.79	812	3.75	1.02
	I carefully planned my course of action to solve the problem. (SR6)	1422	2.91	0.88	811	3.05	1.15
	Before writing the answer, I considered what it requires of me to solve the problem. (SR2)	1431	3.54	0.84	817	3.64	1.11
Monitoring	I checked my work all the way through the problem. (SR5)	1425	3.52	0.89	805	3.62	1.12
	I looked back and checked whether what I had done was right. (SR7)	1437	2.96	1.03	816	2.98	1.34
	I stopped and rethought a step I had already taken. (SR8)	1429	3.38	0.96	809	3.36	1.27
	I looked back at the problem to see if my answer made sense. (SR9)	1423	3.29	1.07	809	3.32	1.49

effort). The test-taking time refers to the time (in minutes) that the student spent on the PT, and it was obtained from the log file of the test session. Recall that the maximum duration of the test was 60 min. The descriptive statistics of test-taking effort, test-taking time as well as the scores of three critical thinking components (analysis and problem-solving, writing effectiveness, and writing) measured in the CLA + test are shown in Table 2.

3.3. Statistical analyses

We applied structural equation modelling (SEM) to examine possible relations between the latent factors of self-regulation, in terms of planning and monitoring, and critical thinking, and the observed variables measuring test-taking effort and time. A graph depicting the postulated structural equation model is presented in Figure 1.

Based on previous research (Bagheri and Ghanizadeh 2016; Bensley et al. 2016; Boekaerts and Cascallar 2006; Ghanizadeh and Mizaee 2012; Hyttinen et al. 2021; Liu et al. 2016; Wise, Kuhfeld, and Soland 2019), the starting point for our modelling was that the intercorrelated factors of planning and monitoring have direct effects on test-taking effort, test-taking time and critical thinking; effort has a direct effect on time and critical thinking; and finally, time has a direct effect on critical thinking. Consequently, planning and monitoring have indirect effects on time (through effort) and critical thinking (through effort and time), and effort has an indirect effect on critical thinking through time.

The target was to find a theoretically reasonable model with adequate fit to the data, while keeping it as parsimonious as possible. Thus, we removed all parameters, which could be considered statistically non-significant, from the model during the model-building process. Given the amount of testing and the size of the data, we used the 0.5 per cent limit ($p < .005$) as the criterion for statistical significance to reduce the risk of excessively liberal inference. However, the data set was not a true random sample from the Finnish student population due to the voluntary participation of institutions, although we could consider it nationally representative due to the stratification by fields of study. Given this, we did not interpret the statistical significances strictly in the conventional sense, that is, as probabilities of false rejections of null hypotheses. Instead, we employed the test statistics more like signal-to-noise ratios, which help in determining which parameters are important in model fitting and which are not.

The goodness-of-fit of the model was assessed by the criteria CFI, SRMR, and RMSEA, which are usually employed in structural equation modelling. In particular, the distribution of test-taking time appeared considerably skewed (the vast majority of the students spent 45 min or more on the test). We therefore estimated the SEM parameters with robust maximum likelihood (MLR) instead of the usual maximum likelihood (ML), as recommended by Muthén and Muthén (2015).

As the data were collected using a stratified cluster design, where the inclusion probabilities were not equal for all students, survey weights were applied in all analyses to correct the possibly resulting distortions in the sample. The weights were derived using a national register containing the population of students in the Finnish higher education institutions. To obtain the weights, the student population was broken down into subpopulations by institution type (university or university of applied sciences), field of study, student class, and sex. The weights were then determined as

Table 2. Descriptive statistics of test-taking effort, test-taking time, and components of critical thinking.

	Initial-stage students			Final-stage students		
	n	Mean	SD	n	Mean	SD
Test-taking effort	1467	3.37	0.70	829	3.53	0.89
Test-taking time (minutes)	1469	49.3	11.6	831	49.4	14.3
Analysis and problem-solving (APS)	1469	3.97	0.67	831	4.16	0.91
Writing effectiveness (WE)	1469	3.97	0.72	831	4.14	0.95
Writing mechanics (WM)	1469	4.23	0.72	831	4.49	0.92

ratios of the subpopulation totals to the corresponding sample totals, scaled to add up to the total sample size of 2402. Then, the weighted distributions of institution type, field of study, class, and sex in the sample agree with those in the population.

Clustering can increase the uncertainty of estimates and it must be taken into account in the statistical analyses (Kish 1965). We handled the clustering by using design-based methods tailored for analysis of complex survey data. The descriptive statistics were calculated by the Surveymeans procedure of SAS® software, Version 9.4, and SEM analyses were carried out by Mplus® software, Version 7, with analysis option Complex.

4. Results

We tested the structural equation model illustrated in Figure 1 separately for initial-stage students ($n = 1335$) and final-stage students ($n = 761$). We treated initial-stage and final-stage students as different groups because their level of critical thinking and self-regulation might differ due to the fact that their study experience in higher education differs. This might further lead to differing associations between the variables of interest. However, the final models for initial-stage and final-stage students appeared similar: only minor differences in the magnitudes of the estimates were observed. The final models are presented graphically in Figures 2 and 3, containing statistically significant parameters only. All parameter estimates given in the Figures are standardised, and their standard errors are given in parentheses. We consider the model for initial-stage students first.

The model fit for initial-stage students was good (RMSEA = 0.04, CFI = 0.96, SRMR = 0.05). The model explained 36 per cent of the variation in critical thinking, 26 per cent of the variation in test-taking time, and 15 per cent of the variation in test-taking effort.

According to the estimated coefficients, the 'main' path goes from planning to effort, from effort to time, and from time to critical thinking. Additionally, planning had a significant direct effect on time, and effort had a significant direct effect on critical thinking. The null role of monitoring is notable. Monitoring only correlated with planning but had no statistically significant effects on other variables in the model. Overall, the results suggest that the link from self-regulation (planning) to critical thinking performance is mainly indirect, mediated by effort and time: planning increases the amount of actual effort, which then manifests itself in increased test-taking time. As the distribution of test-taking time showed remarkable skewness, the interpretation of its effect on critical

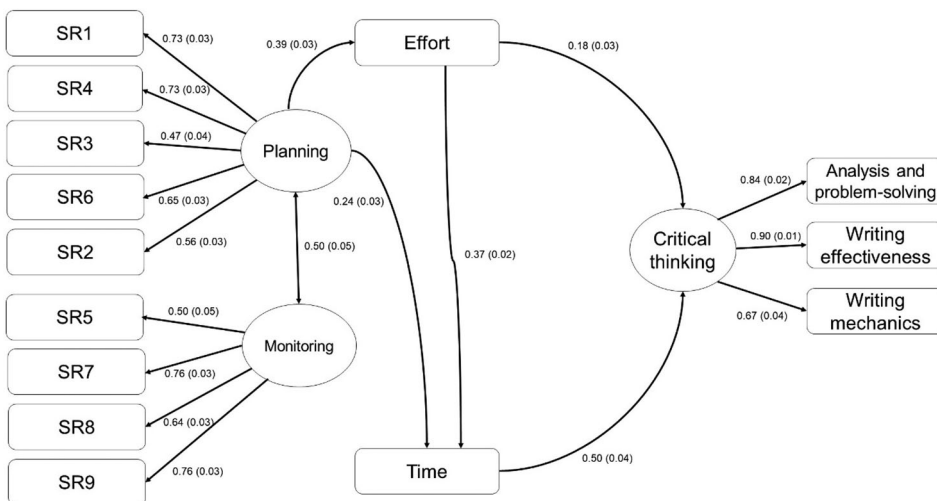


Figure 2. The structural equation model of the relationships between planning, monitoring, effort, time and critical thinking among initial-stage students.

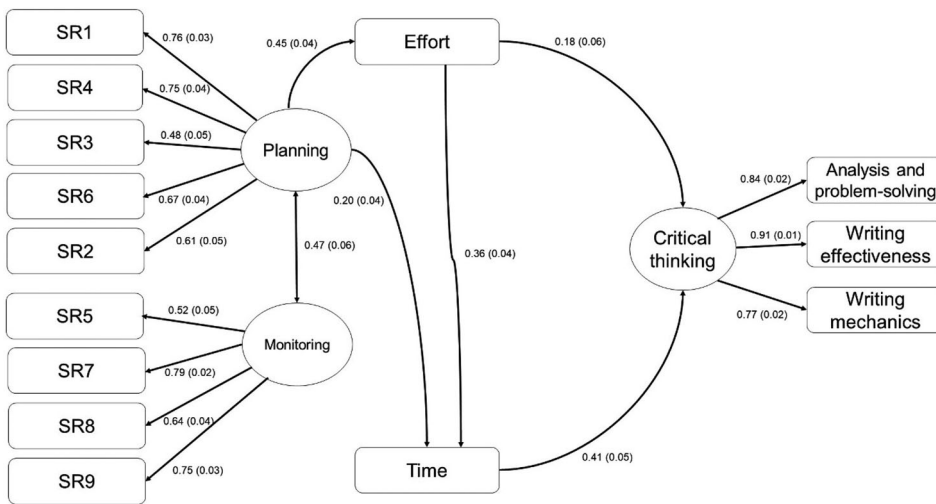


Figure 3. The structural equation model of the relationships between planning, monitoring, effort, time and critical thinking among final-stage students.

thinking is that it was hard to obtain high test scores without spending maximum or almost maximum time in the session. Nevertheless, the spending of maximum time was not a guarantee of high scores.

Table 3 shows the estimated direct, total indirect, and total effects calculated from the model parameters. We particularly underline that, in terms of planning, the effect of self-regulation on critical thinking was completely indirect. All effects given in Table 3 were statistically significant.

Turning to the model for final-stage students (see Figure 3), we note that the goodness-of-fit was adequate (RMSEA = 0.04, CFI = 0.96, SRMR = 0.05). The model explained 27 per cent of the variation in critical thinking, 23 per cent of the variation in test-taking time, and 20 per cent of the variation in test-taking effort.

The model for final-stage students is strikingly equivalent to the model for initial-stage students. The path structure was the same in both models. Even the parameter estimates in the two models did not differ from each other; the only significant difference was that the factor loading of writing mechanics on critical thinking was larger for final-stage students than for initial-stage students.

Table 4 shows the direct, indirect, and total effect estimates calculated from the parameters of the model for final-stage students. We note again that the effect of planning on critical thinking was purely indirect, mediated by effort and time. All effects given in the table were statistically significant.

5. Discussion and conclusions

5.1. Key findings in the light of previous literature

This research contributes to existing knowledge about performance-based assessments of critical thinking by providing new information on how planning and self-monitoring, together with test-

Table 3. Direct, total indirect and total effects in the model for initial-stage students.

	Direct	Total indirect	Total
Planning → Effort	0.39		0.39
Planning → Time	0.24	0.15	0.39
Planning → Critical thinking	0.37	0.26	0.26
Effort → Time	0.18	0.18	0.37
Effort → Critical thinking	0.50		0.36
Time → Critical thinking			0.50

Table 4. Direct, total indirect and total effects in the model for final-stage students.

	Direct	Total indirect	Total
Planning → Effort	0.45		0.45
Planning → Time	0.20	0.16	0.36
Planning → Critical thinking	0.36	0.23	0.23
Effort → Time	0.18	0.15	0.36
Effort → Critical thinking	0.41		0.33
Time → Critical thinking			0.41

taking effort and time, relate to students' critical thinking performance. Based on earlier studies, we expected that both self-regulation (Bagheri and Ghanizadeh 2016; Bensley et al. 2016; Boekaerts and Cascallar 2006; Ghanizadeh and Mizae 2012) and test-taking effort and time (Liu et al. 2016; Silm, Must, and Täht 2019; Silm, Pedaste, and Täht 2020; Wise, Kuhfeld, and Soland 2019) contribute to higher education students' critical thinking performance. Surprisingly, the findings of this study indicated that the contribution of self-regulation in test-taking to undergraduate students' critical thinking performance was limited and interestingly mediated by other factors indicating the complexity and situationality of the process. We found that planning has a role in critical thinking performance. However, in contrast to earlier findings (Bagheri and Ghanizadeh 2016; Ghanizadeh and Mizae 2012), no evidence of the connection between monitoring and critical thinking was detected in the context of performance-based assessment. Moreover, the relation between planning and critical thinking test performance was indirect: planning manifests in increased effort and time, which then associates with higher performance task scores. In this study, test-taking time had the strongest impact on students' performance task scores. The model was tested in two groups of higher education students. The associations did not differ between the initial-stage and final-stage students, indicating that years in undergraduate studies do not seem to be an important factor explaining the associations with critical thinking performance, test-taking effort, and self-regulation in this specific context of critical thinking assessment.

There could be several reasons for the limited role of self-regulation in critical thinking performance. One reason could be the fact that the specific performance task used in this study was relatively straightforward (Nissinen et al. 2021). Thus, if the assessment task had been more demanding, it would possibly have required more monitoring (cf. Hoyle and Dent 2018; Uzuntir-yaki-Kondakci and Çapa-Aydin 2013; Winne 2018). Another reason for the limited connection between self-regulation and critical thinking might be that the time pressure prevented students from using their self-regulation capacity to the full. The finding could also be explained by the self-report method used to measure self-regulation; students might not be able to reflect on their self-regulation accurately (cf. Karabenick et al. 2007). However, it is noteworthy that two scales of self-regulation (i.e. planning and self-monitoring exercised when working on the PT) functioned very well. Another possible explanation for the difference between earlier research and this study might be that previous research on critical thinking has focused on students' general self-regulation of learning or emotions, but not on self-regulation applied in a complex critical thinking assessment (see Bagheri and Ghanizadeh 2016; Ghanizadeh and Mizae 2012). Therefore, the impact of the components of self-regulation on students' critical thinking needs to be further investigated. Research applying a longitudinal research design should be undertaken to investigate changes in critical thinking and their relations to self-regulation during higher education studies. Furthermore, disciplinary differences should be studied when exploring the changes in the association between critical thinking and self-regulation.

Although this study focuses on one particular performance task, the results in relation to self-regulation and effort in test-taking have wider relevance in research on critical thinking. The results confirm that responding successfully to performance-based critical thinking assessments involves not only certain skills, but also considerable effort in using these skills to meet the demands of the task (Bensley et al. 2016; Facione 1990; Halpern 2014; Hyytinen et al. 2021).

Additionally, the results of this study showed that students' test-taking effort and time were strong predictors of critical thinking performance. This also accords with earlier research which has shown that test-taking effort and time have a substantial impact on critical thinking performance (Bensley et al. 2016; Liu et al. 2016). The findings reported here shed new light on the association between self-regulation, test-taking effort, and time. The findings indicated that together with test-taking effort and time, planning influences students' critical thinking performance. Thus, when interpreting undergraduate students' level of critical thinking using performance-based assessment, it is important to take into account the effort and time expended by students, especially in low-stakes testing (Silm, Must, and Täht 2019; Silm, Pedaste, and Täht 2020).

5.2. Limitations of the study and future research

This study has several limitations. The first of these relates to the nature of the sample. As participating in the study was voluntary for the higher education institutions, the collected data cannot be considered a truly random sample from the Finnish student population. However, by using stratification by field of study in the data collection together with survey weighting, we can assume that the student cohorts in question are represented reasonably well in the data, across the spectrum of fields provided in the sector of Finnish higher education. Second, the study was conducted in the context of one specific performance-based assessment task. Therefore, in future it would be important to study the interconnection between critical thinking, test-taking effort, and self-regulation in the different contexts of performance tasks. A third limitation is that the survey responses on self-regulation in test-taking were based on self-report. Hence, it would be important to develop performance tasks to measure students' self-regulation skills.

The findings provide insights into enhancing performance-based critical thinking assessments. To better capture the complex nature of critical thinking, assessment tasks should trigger the multifaceted construct of critical thinking accordingly (Kane, Crooks, and Cohen 2005; McClelland 1973; Shavelson et al. 2019). It follows that performance-based assessments should be carefully designed and implemented so that they really trigger the construct of critical thinking, including a set of skills and affective dispositions together with self-regulation.

5.3. Practical implications

The findings of the study have several practical implications for the higher education context, especially regarding support for the learning of critical thinking and self-regulation skills. This is particularly significant since these are essential skills for students to learn during higher education (Arum and Roksa 2011; Schunk and Greene 2018). It is important that teachers in higher education are aware of the characteristics of these processes and are able to support their development in a relevant way through their teaching and pedagogical practices. As the results reflect, these are particularly complex processes and skills, and it is likely that several contextual, situational and individual factors have an influence on them (cf. Bensley et al. 2016; Rear 2019). Thus, students need to be given a variety of opportunities to practise critical thinking and self-regulation skills throughout their studies (Hyytinen et al. 2021). It is also essential to discuss these processes and skills as well as their development with students and to make their monitoring and evaluation explicit (cf. Halpern 2014). The use of self-evaluations and versatile feedback from peers and teachers throughout the learning processes may support the acquisition of critical thinking and self-regulation skills.

Acknowledgements

The authors would like to thank the Council for Aid to Education (CAE) for providing the CLA + International. The funder and the CAE had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the Finnish Ministry of Education and Culture. The study was carried out through the funding of a project entitled *KAPPAS! – Korkeakouluopiskelijoiden oppimistulosten arviointi Suomessa* (Assessment of Undergraduate Students' Learning Outcomes).

ORCID

Heidi Hyytinen  <http://orcid.org/0000-0002-5078-1429>

Kari Nissinen  <http://orcid.org/0000-0001-5064-2552>

Katri Kleemola  <http://orcid.org/0000-0002-9274-9718>

Jani Ursin  <http://orcid.org/0000-0001-6557-9058>

Auli Toom  <http://orcid.org/0000-0002-3261-3376>

References

- Aloisi, C., and A. Callaghan. 2018. "Threats to the Validity of the Collegiate Learning Assessment (CLA+) as a Measure of Critical Thinking Skills and Implications for Learning Gain." *Higher Education Pedagogies* 3 (1): 57–82. <https://doi.org/10.1080/23752696.2018.1449128>.
- Arum, R., and J. Roksa. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago, IL: University of Chicago Press.
- Badcock, P., P. Pattison, and K. L. Harris. 2010. "Developing Generic Skills Through University Study: A Study of Arts, Science and Engineering in Australia." *Higher Education* 60 (4): 441–58. <https://doi.org/10.1007/s10734-010-9308-8>.
- Bagheri, F., and A. Ghanizadeh. 2016. "Critical Thinking and Gender Differences in Academic Self-Regulation in Higher Education." *Journal of Applied Linguistics and Language Research* 3 (3): 133–45.
- Bensley, D., C. Rainey, M. P. Murtagh, J. Flinn, C. Maschiochi, P. C. Bernhardt, and S. Kuehne. 2016. "Closing the Assessment Loop on Critical Thinking: The Challenges of Multidimensional Testing and Low Test-Taking Motivation." *Thinking Skills and Creativity* 21 (September): 158–68. <https://doi.org/10.1016/j.tsc.2016.06.006>.
- Boekaerts, M., and E. Cascallar. 2006. "How Far Have We Moved Toward the Integration of Theory and Practice in Self-Regulation?" *Educational Psychology Review* 18 (3): 199–210. <https://doi.org/10.1007/s10648-006-9013-4>.
- Braun, H. 2019. "Performance Assessment and Standardization in Higher Education: A Problematic Conjunction?" *British Journal of Educational Psychology* 89 (3): 429–40. <https://doi.org/10.1111/bjep.12274>.
- Braun, H., R. J. Shavelson, O. Zlatkin-Troitschanskaia, and K. Borowiec. 2020. "Performance Assessment of Critical Thinking: Conceptualization, Design, and Implementation." *Frontiers in Education* 5: 156. <https://doi.org/10.3389/educ.2020.00156>.
- Davey, T., S. Ferrara, P. W. Holland, R. J. Shavelson, N. M. Webb, and L. Wise. 2015. *Psychometric Considerations for the Next Generation of Performance Assessment*. Washington, DC: Center for K-12 Assessment & Performance Management, Educational Testing Service. https://www.ets.org/Media/Research/pdf/psychometric_considerations_white_paper.pdf.
- Ennis, R. H. 1996. "Critical Thinking Dispositions: Their Nature and Assessability." *Informal Logic* 18 (2), <https://doi.org/10.22329/il.v18i2.2378>.
- Ercikan, K., and M. E. Oliveri. 2016. "In Search of Validity Evidence in Support of the Interpretation and Use of Assessments of Complex Constructs: Discussion of Research on Assessing 21st Century Skills." *Applied Measurement in Education* 29 (4): 310–8. <https://doi.org/10.1080/08957347.2016.1209210>.
- Evens, M., A. Verburgh, and J. Elen. 2013. "Critical Thinking in College Freshmen: The Impact of Secondary and Higher Education." *International Journal of Higher Education* 2 (3): 139. <https://doi.org/10.5430/ijhe.v2n3p139>.
- Facione, P. A. 1990. *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. The Delphi Report. Millbrae, CA: California Academic Press.
- Ghanizadeh, A., and S. Mizaee. 2012. "'EFL Learners' Self-Regulation, Critical Thinking and Language Achievement." *International Journal of Linguistics* 4 (3): 444–61. <https://doi.org/10.5296/ijl.v4i3.1979>.
- Gurcay, D., and H. O. Ferah. 2018. "'High School Students' Critical Thinking Related to Their Metacognitive Self-Regulation and Physics Self-Efficacy Beliefs." *Journal of Education and Training Studies* 6 (4): 125. <https://doi.org/10.11114/jets.v6i4.2980>.
- Halpern, D. F. 2014. *Thought and Knowledge: An Introduction to Critical Thinking*. 5th ed. New York, NY: Psychology Press.

- Holma, K. 2016. "The Critical Spirit: Emotional and Moral Dimensions Critical Thinking." *Studier i Pædagogisk Filosofi* 4 (1): 17. <https://doi.org/10.7146/spf.v4i1.18280>.
- Hoyle, R. H., and A. L. Dent. 2018. "Developmental Trajectories of Skills and Abilities Relevant for Self-Regulation of Learning and Performance." In *Handbook of Self-Regulation of Learning and Performance*, edited by Dale H. Schunk and Jeffrey Alan Greene, 2nd ed., 49–63. *Educational Psychology Handbook Series*. New York, NY: Routledge, Taylor & Francis Group.
- Hu, L., and P. M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1): 1–55. <https://doi.org/10.1080/10705519909540118>.
- Hyytinen, H., J. Ursin, K. Silvennoinen, K. Kleemola, and A. Toom. 2021. "The Dynamic Relationship between Response Processes and Self-Regulation in Critical Thinking Assessments." *Studies in Educational Evaluation* 71 (December): 101090. <https://doi.org/10.1016/j.stueduc.2021.101090>.
- Kane, M., T. Crooks, and A. Cohen. 2005. "Validating Measures of Performance." *Educational Measurement: Issues and Practice* 18 (2): 5–17. <https://doi.org/10.1111/j.1745-3992.1999.tb00010.x>.
- Karabenick, S. A., M. E. Woolley, J. M. Friedel, B. V. Ammon, J. Blazeovski, C. R. Bonney, E. De Groot, et al. 2007. "Cognitive Processing of Self-Report Items in Educational Research: Do They Think What We Mean?" *Educational Psychologist* 42 (3): 139–51. <https://doi.org/10.1080/00461520701416231>.
- Kish, L. 1965. *Survey Sampling*. New York, NY: John Wiley & Sons.
- Kleemola, K., H. Hyytinen, and A. Toom. 2022a. "Exploring Internal Structure of a Performance-Based Critical Thinking Assessment for New Students in Higher Education." *Assessment & Evaluation in Higher Education* 47 (4): 556–69. <https://doi.org/10.1080/02602938.2021.1946482>.
- Kleemola, K., H. Hyytinen, and A. Toom. 2022b. "Critical Thinking and Writing in Transition to Higher Education in Finland: Do Prior Academic Performance and Socioeconomic Background Matter?" *European Journal of Higher Education*, 1–21. <https://doi.org/10.1080/21568235.2022.2075417>.
- Kuhn, D. 2019. "Critical Thinking as Discourse." *Human Development* 62 (3): 146–64. <https://doi.org/10.1159/000500171>.
- Liu, O. L., L. Mao, L. Frankel, and J. Xu. 2016. "Assessing Critical Thinking in Higher Education: The HELghten™ Approach and Preliminary Validity Evidence." *Assessment & Evaluation in Higher Education* 41 (5): 677–94. <https://doi.org/10.1080/02602938.2016.1168358>.
- Maksum, A., I. W. Widiana, and A. Marini. 2021. "Path Analysis of Self-Regulation, Social Skills, Critical Thinking and Problem-Solving Ability on Social Studies Learning Outcomes." *International Journal of Instruction* 14 (3): 613–28. <https://doi.org/10.29333/iji.2021.14336a>
- McClelland, D. C. 1973. "Testing for Competence Rather than for 'Intelligence.'" *American Psychologist* 28 (1): 1–14. <https://doi.org/10.1037/h0034092>.
- Muthén, L. K., and B. O. Muthén. 2015. *Mplus User's Guide*. 7th ed. Los Angeles, CA: Muthén & Muthén.
- Nissinen, K., J. Ursin, H. Hyytinen, and K. Kleemola. 2021. "Higher Education Students' Generic Skills." In *Assessment of Undergraduate Students' Generic Skills – Findings of the Kappas! Project*, edited by J. Ursin, H. Hyytinen, and K. Silvennoinen. 6 vols., 38–79. Helsinki: Ministry of Education and Culture.
- Räsänen, M., L. Postareff, and S. Lindblom-Ylänne. 2016. "University Students' Self- and Co-Regulation of Learning and Processes of Understanding: A Person-Oriented Approach." *Learning and Individual Differences* 47 (April): 281–8. <https://doi.org/10.1016/j.lindif.2016.01.006>.
- Rear, D. 2019. "One Size Fits All? The Limitations of Standardised Assessment in Critical Thinking." *Assessment & Evaluation in Higher Education* 44 (5): 664–75. <https://doi.org/10.1080/02602938.2018.1526255>.
- Saariaho, E., A. Toom, T. Soini, J. Pietarinen, and K. Pyhäältö. 2019. "Student Teachers' and Pupils' Co-Regulated Learning Behaviours in Authentic Classroom Situations in Teaching Practicums." *Teaching and Teacher Education* 85 (October): 92–104. <https://doi.org/10.1016/j.tate.2019.06.003>.
- Schunk, D. H., and J. A. Greene. 2018. "Historical, Contemporary, and Future Perspectives on Self-Regulated Learning and Performance." In *Handbook of Self-Regulation of Learning and Performance*, edited by Dale H. Schunk and Jeffrey Alan Greene, 2nd ed., 1–16. *Educational Psychology Handbook Series*. NY, New York: Routledge, Taylor & Francis Group.
- Shavelson, R. J., O. Zlatkin-Troitschanskaia, K. Beck, S. Schmidt, and J. P. Marino. 2019. "Assessment of University Students' Critical Thinking: Next Generation Performance Assessment." *International Journal of Testing* 19 (4): 337–62. <https://doi.org/10.1080/15305058.2018.1543309>.
- Silm, G., O. Must, and K. Täht. 2019. "Predicting Performance in a Low-Stakes Test Using Self-Reported and Time-Based Measures of Effort." *Trames Journal of the Humanities and Social Sciences* 23 (3): 353. <https://doi.org/10.3176/tr.2019.3.06>.
- Silm, G., M. Pedaste, and K. Täht. 2020. "The Relationship between Performance and Test-Taking Effort When Measured with Self-Report or Time-Based Instruments: A Meta-Analytic Review." *Educational Research Review* 31 (November): 100335. <https://doi.org/10.1016/j.edurev.2020.100335>.
- Toering, T., M. T. Elferink-Gemser, L. Jonker, M. J. G. van Heuvelen, and C. Visscher. 2012. "Measuring Self-Regulation in a Learning Context: Reliability and Validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS)." *International Journal of Sport and Exercise Psychology* 10 (1): 24–38. <https://doi.org/10.1080/1612197X.2012.645132>.

- Tremblay, K., D. Lalancette, and D. Roseveare. 2012. *AHELO, Assessment of Higher Education Learning Outcomes. Feasibility Study Report, Vol. 1. Design and Implementation. Study Report*. Paris: OECD. <https://www.oecd.org/education/skills-beyond-school/AHELOFSReportVolume1.pdf>.
- Ursin, J. 2020. "Assessment of Higher Education Learning Outcomes Feasibility Study." In *The SAGE Encyclopedia of Higher Education*, edited by M. David and M. Amey, 132–3. Thousand Oaks, CA: SAGE Publications, Inc.
- Ursin, J., and H. Hyytinen. 2022. "Assessing the Generic Skills of Undergraduate Students in Finland." In *Does Higher Education Teach Students to Think Critically?*, edited by D. Van Damme and D. Zahner, 179–96. Paris: OECD Publishing.
- Usher, E. L., and D. H. Schunk. 2018. "Social Cognitive Theoretical Perspective of Self-Regulation." In *Handbook of Self-Regulation of Learning and Performance*, edited by D. H. Schunk and J. A. Greene, 2nd ed., 17–35. *Educational Psychology Handbook Series*. New York, NY: Routledge, Taylor & Francis Group.
- Utriainen, J., M. Marttunen, E. Kallio, and P. Tynjälä. 2017. "University Applicants' Critical Thinking Skills: The Case of the Finnish Educational Sciences." *Scandinavian Journal of Educational Research* 61 (6): 629–49. <https://doi.org/10.1080/00313831.2016.1173092>.
- Uzuntiryaki-Kondakci, E., and Y. Çapa-Aydin. 2013. "Predicting Critical Thinking Skills of University Students Through Metacognitive Self-Regulation Skills and Chemistry Self-Efficacy." *Educational Sciences: Theory & Practice* 13 (1): 666–70.
- Winne, P. H. 2018. "Cognition and Metacognition Within Self-Regulated learning." In *Handbook of Self-Regulation of Learning and Performance*, edited by D. H. Schunk and J. A. Greene, 2nd ed., 36–48. *Educational Psychology Handbook Series*. New York, NY: Routledge, Taylor & Francis Group.
- Wise, S. L., M. R. Kuhfeld, and J. Soland. 2019. "The Effects of Effort Monitoring With Proctor Notification on Test-Taking Engagement, Test Performance, and Validity." *Applied Measurement in Education* 32 (2): 183–92. <https://doi.org/10.1080/08957347.2019.1577248>.
- Zahner, D., and A. Ciolfi. 2018. "International Comparison of a Performance-Based Assessment in Higher Education." In *Assessment of Learning Outcomes in Higher Education: Cross-National Comparisons and Perspectives*, edited by O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, C. Lautenbach, and C. Kuhn, 215–44. *Methodology of Educational Measurement and Assessment*. Cham: Springer International Publishing.
- Zimmerman, B. J. 2002. "Becoming a Self-Regulated Learner: An Overview." *Theory Into Practice* 41 (2): 64–70. https://doi.org/10.1207/s15430421tip4102_2.
- Zlatkin-Troitschanskaia, O., R. J. Shavelson, and C. Kuhn. 2015. "The International State of Research on Measurement of Competency in Higher Education." *Studies in Higher Education* 40 (3): 393–411. doi:10.1080/03075079.2015.1004241.