



Exploring First Semester Changes in Domain-Specific Critical Thinking

Tine Nielsen^{1,2*}, Inmaculada Martínez-García³ and Enrique Alastor³

¹ Department of Applied Research in Education and Social Sciences, UCL University College [UCL Erhvervsakademi og Professionshøjskole], Odense, Denmark, ² Department of Psychology, University of Copenhagen, Copenhagen, Denmark, ³ Department of Didactics and School Organization, Faculty of Educational Sciences, University of Málaga, Málaga, Spain

OPEN ACCESS

Edited by:

Edith Braun,
Justus-Liebig Universität, Germany

Reviewed by:

Kristina Walz,
University of Giessen, Germany
Kari Nissinen,
University of Jyväskylä, Finland

*Correspondence:

Tine Nielsen
tini@ucl.dk

Specialty section:

This article was submitted to
Higher Education,
a section of the journal
Frontiers in Education

Received: 26 February 2022

Accepted: 23 May 2022

Published: 24 June 2022

Citation:

Nielsen T, Martínez-García I and
Alastor E (2022) Exploring First
Semester Changes
in Domain-Specific Critical Thinking.
Front. Educ. 7:884635.
doi: 10.3389/educ.2022.884635

Critical thinking is a common aim for higher education students, often described as general competencies to be acquired through entire programs as well as domain-specific skills to be acquired within subjects. The aim of the study was to investigate whether statistics-specific critical thinking changed from the start of the first semester to the start of the second semester of a two-semester statistics course, where the curriculum contains learning objectives and assessment criteria related to critical thinking. The brief version of the Critical Thinking scale (CTh) from the Motivated Strategies of Learning Questionnaire addresses the core aspects of critical thinking common to three different definitions of critical thinking. Students rate item statements in relation to their statistics course using a frequency scale: 1 = never, 2 = rarely, 3 = sometimes, 4 = often, and 5 = always. Participants were two consecutive year-cohorts of full-time Bachelor of Psychology students taking a two-semester long statistics course placed in the first two semesters. Data were collected in class with a paper-pencil survey 1 month into their first semester and again 1 month into the second. The study sample consisted of 336 students ($n_{\text{cohort 1}} = 166$, $n_{\text{cohort 2}} = 170$) at baseline, the follow-up was completed by 270 students with 165 students who could be matched to their baseline response. To investigate the measurement properties of the CTh scale, item analysis by the Rasch model was conducted on baseline data and subsequently on follow-up data. Change scores at the group level were calculated as the standardized effect size (ES) (i.e., the difference between baseline and follow-up scores relative to the standard deviation of the baseline scores). Data fitted Rasch models at baseline and follow-up. The targeting of the CTh scale to the student sample was excellent at both timepoints. Absolute individual changes on the CTh ranged from -5.3 to 5.1 points, thus showing large individual changes in critical thinking. The overall standardized effect was small and negative (-0.12), with some variation in student strata defined by, gender, age, perceived adequacy of math knowledge to learn statistics, and expectation to need statistics in future employment.

Keywords: critical thinking, domain-specific, changes, higher education, Rasch model, statistics in psychology

INTRODUCTION

Critical thinking is a central concept in higher education, and as it has become relevant at both the individual and societal level it will not only improve students' academic success but also the quality of education (Ren et al., 2020). The scientific literature has investigated the responsibility of educational institutions in training students in competencies that enables them to be future citizens ready to be an acting part of the society by making them critical thinkers (Kuhn, 1999; Paul and Elder, 2005). Thus, there has been a growing interest in the incorporation of critical thinking in the education curricula making critical thinking one of the main aims (Lau, 2015; McQuirk, 2021). With regard to the outcome of higher education, critical thinking is predominantly construed as generic, as it is described in terms of the competencies, students are expected to possess at the completion of a degree program (see **Supplementary Appendix 2** for the competency description for a degree program in this study). However, in terms of incorporating critical thinking into higher education programs, this appears rarely to be in the form of independent courses teaching critical thinking. More often critical thinking seems to be implemented through teaching methods and specifically designed activities within subject courses thus construing critical thinking as domain-specific, or simply by using the term critical thinking in the curriculum description without clear definitions, program- or course-determined approaches to teaching toward critical thinking (c.f. **Supplementary Appendix 2** for the current study). These two levels of implementing critical thinking in higher education tie to the discussion of critical thinking as generic/general or domain- and subject-specific.

There are different ways of understanding critical thinking that involve different implications for practice, so there is no consensus on a single definition (Moseley et al., 2005). Commonly in the literature, there is a distinction between thinking or cognitive skills and dispositional aspects of critical thinking, but as two sides of critical thinking and not separate positions. Two prevalent authors in the field, whose definitions or instruments many draw on in their research, are Facione and Halpern. Facione (1990) conducted a large Delphi study to narrow down the components of critical thinking, and the panel reached a consensus conceptualization of critical thinking as consisting of two dimensions: cognitive skills and affective dispositions. He further defines critical thinking as “*the purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based*” (Facione, 1990, p. 3). Halpern (2014) understand critical thinking as “*the deliberate use of skills and strategies that increase the probability of a desirable outcome*” (p. 450) and that critical thinking is involved in “*solving problems, formulating inferences, calculating likelihoods, and making decisions*” (p. 8), and thus also refer to both skills and dispositions. Facione and Halpern also make the distinction of critical thinking skills being assessed as the abilities to demonstrate critical skills in tasks or assignments, while critical

thinking dispositions are assessed by self-report instruments. However, in the empirical studies in the field, there is no consensus on this. Thus, studies using self-report instruments have claimed to assess critical thinking skills (e.g., Ricketts and Rudd, 2005), studies employing critical thinking dispositions self-report instrument has claimed to assess critical thinking skills with this (e.g., Kanbay et al., 2017), and lastly, studies claiming to assess abilities are doing this through to some degree subjective teacher evaluations using short rubrics¹ (e.g., Ralston and Bays, 2015). See the following sections for more details on these studies. At a general level, there appears to be a conceptual shift toward using the term skills and then differentiating between assessed and self-report. Thus, in the remainder of this article, we simply use the term critical thinking skills, while recognizing that we use a self-report instrument to assess this, thus assessing students' perceptions of their critical thinking skills.

Critical Thinking as Generic/General Skills or Domain-Specific Skills

One particularly pertinent discussion in the field is whether critical thinking skills are generic/general skills or whether they are domain-/subject-specific (Tiruneh et al., 2017).

The view of critical thinking as a generic set of skills applicable across domains is based on the common features of critical thinking tasks across a wide variety of domains (e.g., Halpern, 1998; Kuhn, 1999). While Halpern (1998) is a proponent of critical thinking as a set of generic skills, her “Four-Part Model for Enhancing Critical Thinking” to teach critical thinking acknowledges that critical thinking takes place within a knowledge domain and should be taught within this domain. However, this does not mean that Halpern considers critical thinking as domain-specific, but rather that the domain is the learning context for skills, which can be applied more universally across domains after being mastered. The view of critical thinking as domain-specific emphasizes that different domains have different criteria relating to critical thinking and thus the skills required inevitably vary across domains (e.g., McPeck, 1992; Moore, 2011). The issue is more likely not an either/or issue, but an issue of both in combination, as content and critical thinking tasks and skills might differ across domains as they are invariably linked to the domain-knowledge, but there are also commonalities across domains, due to the cognitive processes involved in critical thinking (e.g., Bailin et al., 1999). As such, critical thinking may be regarded as a set of domain-specific skills of which some also belong to the set of generic or general critical thinking skills. Whether there is in fact a transfer effect from the domain-specific learning of critical thinking skills to other domains or adding on to generic critical thinking skills, as suggested in some of the literature, is another pertinent issue in the critical thinking field. However, this is not a central topic in the current study, as we are concerned with domain-specific

¹Subjective in the sense that the rubrics have so few categories and the descriptions of categories are so general that even an identical scorings can ensure that the behavior or products rated by the teachers as indicating critical thinking is the same (c.f. Ralston and Bays, 2015).

critical thinking skills and their development in the first part of university studies.

Critical Thinking Skills in First-Year University Students and Their Development

First-year university students are particularly interesting when it comes to studying critical thinking skills and how they develop, as many higher education teachers and researchers concur that “first-year students often enter higher education without the ability to use higher-order thinking skills to master their studies” (De Jager, 2012, p. 1374). Much of the research into critical thinking skills of first-year university students and the development of critical thinking skills during university, has been focused on the development of teaching models and methods to enhance critical thinking, assessing their effects, and comparing how different teaching methods affect the critical thinking of the students. One example is Saenab et al. (2021) who developed the ReCODE model (Reading, Connecting, Observing, Discussing, and Evaluating) to improve first-year Biology students’ acquisition of critical thinking. The outcome was positive with regard to enhancing students’ critical thinking over the course of 3 months, however, it was only used on 38 students. Thomas (2011) developed the “Embedding generic skills in a business curriculum”-program consisting of activities and assessment resources for university teachers to develop critical thinking skills with their first-year students, and emphasize that these skills should be developed in the first year. The suggestions were not tested. On a similar note, Hammer and Green (2011) redesigned a written assessment in the form of a case-based business report for first-year management students in order to facilitate better development of critical thinking as this was part of the requirements for passing. The authors used the percentage of passing students to evaluate the success of the redesign – this went from 78.8 to 84% – but details of the teachers’ assessments were not provided, and thus how critical thinking was assessed was not divulged beyond its being a teacher assessment. Ralston and Bays (2015), on the other hand, found that Engineering students’ ($n = 182$) critical thinking increased during the course of their undergraduate studies, which had purposely been designed to incorporate assignments focused on critical thinking. A four-point, holistic critical thinking rubric was designed for the purpose of the study to evaluate domain-specific critical thinking. As a final example, Tiruneh et al. (2017) compared both domain-specific and general critical thinking skills for first-year students in an introductory Physics course ($n = 143$), using the Halpern Critical Thinking Assessment (HCTA; Halpern, 2015); a standardized scenario-based instrument with 25 everyday scenarios assessing general critical thinking skills by means of computerized scoring in combination with trained grader scoring. The study compared three different instructional designs and found that students in what they termed immersion and infusion designs (intervention) outperformed students in the control design significantly with regard to domain-specific critical thinking as well as course achievement. However, neither

of the intervention designs fostered the acquisition of general critical thinking skills.

It is evident that there is an abundance of studies on various methods to enhance students’ critical thinking skills in the first year and over the course of university studies. However, critical thinking in first-year students has also been investigated with regard to its “natural” development over time (i.e., no particular design implemented to enhance critical thinking) and how critical thinking is related to other psychological and educational constructs, e.g., emotional intelligence (Kaya et al., 2017; Sahanowas and Halder, 2020) and perceived academic control (Stupnisky et al., 2008). Sahanowas and Halder (2020) used the University of Florida - Engagement, Cognitive Maturity and Innovativeness assessment (UF-EMI, Ricketts and Rudd, 2005), which is a self-report instrument measuring generic critical thinking, in a cross-sectional study with the first-year students in various disciplines ($n = 500$) found that emotional intelligence was positively related to critical thinking. Kaya et al. (2017) in their study of Nursing students find that they possess a low level of critical thinking at the start of the first academic year, and while critical thinking was positively associated with emotional intelligence at the start, neither developed over the course of the year. Kaya et al. (2017) made use of a Turkish translation of the California Critical Thinking Disposition Scale (Facione et al., 1998), which is a self-report measure of generic critical thinking. Stupnisky et al. (2008) conducted a longitudinal study with Psychology students ($n = 1,196$) with the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991), which contains a domain-specific self-report critical thinking scale, and found a reciprocal relationship between critical thinking and perceived academic control, so that students perceived academic control 1 month into the first year predicted critical thinking 6 months later, while critical thinking 1 month into the first year also predicted perceived academic control 6 months later. Another example, of a study on the “natural” development of critical thinking over time, is Kanbay et al. (2017) who assessed critical thinking in Nursing students ($n = 46$), with the (California Critical Thinking Disposition Scale, see above) at the start of the first year and at the end of the second, third and fourth years of study. Their results revealed a medium level of critical thinking at the beginning and no improvement in critical thinking across the four-year period of time, not statistically and not at the absolute level. In a qualitative study, Özelçi and Çalışkan, 2019, interviewed 11 teacher candidates two times about their critical thinking. The results showed no change in self-perception in critical thinking from the first to the fourth year of study.

Development of Students’ Statistics-Related Critical Thinking

Turning to the domain-specific concept of statistics-related critical thinking, several studies have been conducted. Bensley et al. (2010) studied the acquisition of critical thinking skills in instructional different groups of students enrolled in a research methods course by the psychological Critical Thinking Test

(Bensley and Baxter, 2006), which is a domain-specific multiple-choice test. More specifically they compared the acquisition of critical thinking skills for analyses of psychological arguments students who had critical thinking skills infused directly into their course with students where this was not the case. The infusing of critical thinking skills consisted of using a methodologically oriented textbook as well as a critical thinking textbook, as well as examples and practice of critical thinking through exercises and corrective feedback. The non-infusing courses used another textbook that embedded statics instruction within a research design and methodology discussion. The study found that the group of students who had received instruction aimed explicitly at critical thinking showed significantly greater gains in argument analysis skills than the students who had received no explicit critical thinking instruction. Contrary to this, Goode et al. (2018) compared how critical thinking was expressed in early and late writing assignments using specific critical thinking learning objectives recommended by the American Psychological Association (i.e., effective use of critical thinking, use of reasoning in argument, and problem-solving effectiveness) for psychology students assigned at random to a face-to-face and a blended learning versions of a statistics and research design course. Goode et al. (2018) developed a domain-specific scoring rubric with three areas being scored from 'does not meet expectations' to 'far exceeds expectations' for the teachers' assessment of critical thinking. The difference between the two instructional designs was simply that the blended learning version of the course was taught as a 50/50 flipped hybrid of the face-to-face course. Thus, in the blended learning hybrid, students attended face-to-face classes once a week rather than two, and for the second weekly class, they viewed online lectures and worked with other materials outside of the class setting. There was no significant difference in the development of critical thinking between students in the face-to-face and students in the blended learning design. However, an instructor effect was found, showing that student assigned to classes by two instructors increased their critical thinking significantly more than students assigned to two other instructors, and for one instructor both randomly assigned groups of students had a decline in critical thinking during the course. Setambah et al. (2019) evaluated how the critical thinking skills of teacher preparation students in their second semester developed in a basic statistics course employing Adventure-Based Learning (ABL) compared to a control group not receiving ABL. They found that after 10 weeks there was no significant difference, while there was weak evidence for a difference favoring the experimental groups after a further 8 weeks. Lastly, Cheng et al. (2018) showed how the critical thinking of undergraduate students taking introductory statistics classes within various degree programs increased across semester-long courses incorporating assignments, in-class discussion, and Socratic dialog. Cheng et al. (2018) designed a domain-specific rubric with four dimensions related to critical thinking to be scored by domain-specialists and as well as a student self-report survey to assess students' perceptions of improvement in critical thinking.

With regard to the domain-specific statistics-related critical thinking, there appears to be a lack of studies on the "natural"

development over time, i.e., without implementation of any specific teaching methodology. As Tiruneh et al. (2017) suggest that "*meaningful instruction in every subject domain inherently comprises the development of CT skills, and therefore, proficiency in CT skills can be achieved as students construct knowledge of a subject-matter domain without any explicit emphasis on the teaching of general CT skills during instruction*" (p. 1067), such studies might contribute to the knowledge of the "natural" development of statistics-related critical thinking.

The Current Study

Drawing on the previous research, the present study intends to contribute to the field by studying specifically the development of statistics-related critical thinking in first-year psychology undergraduate students in a Danish university, where no particular emphasis on teaching critical thinking skills is reflected in the curriculum, but rather implicit references are given and critical thinking is mentioned in the assessment criteria (c.f. Moore, 2011). The primary aim of the study is to investigate whether statistics-related critical thinking changes from the start of the first semester to the start of the second semester of a two-semester-long statistics course, where the curriculum contains learning objectives implying critical thinking and assessment criteria explicitly requiring critical thinking.

At the overall level, we expected all students to have an increase in critical thinking, based on the general goal and performance orientation of these students² in combination with the implicit mention in the learning objectives and particularly the explicit mention of critical thinking in the assessment criteria for the first semester of the course. However, such an overall change might mask differentiated subgroup changes, and subgroup changes in opposite directions might also result in no change at the overall level. With regard to subgroups, we expected that the overall change in critical thinking would differ for subgroups of students dependent on their baseline perception of the adequacy of their own mathematical knowledge for learning statistics as well as their expectation to need statistics in their future employment, as would the students' baseline level of critical thinking. Specifically, we expected:

Students who perceived their mathematical knowledge to be inadequate for learning statistics were less inclined toward critical thinking at baseline compared to students who perceived they had an adequate level of mathematical knowledge, due to their lack of insight into the field. We had no set expectation with regard to the direction of difference in the change in critical thinking dependent on the perception of the adequacy of mathematical knowledge, as this could go both ways. For some students, a perceived lack in the prerequisite knowledge required would be a motivating factor making them engage more and thus possibly increase more in critical thinking compared to the other student group. But the opposite is also likely for some students, i.e., perceiving a lack of prerequisite knowledge might be further dis-engaging leading to a smaller increase in critical thinking

²In Denmark, psychology is one of the top-10 most difficult higher education degree program to be admitted to, as there are a limited number of vacancies to compete for, and thus it requires almost perfect grades to be admitted.

or even a decrease for this subgroup. In addition, students perceiving adequacy in prerequisite knowledge could be expected to engage more due to their insight and thus increase more in critical thinking than students perceiving their pre-requisite knowledge as inadequate.

Students who did not believe they would need statistics in their future employment were less inclined toward critical thinking at baseline and compared to students believing they would be needing statistics, as they would not be as likely to engage in the cognitively demanding critical because it would be perceived as unnecessary. We would not expect that students' beliefs about their future employment to change much over the course of the first semester of study, and thus we expected that the largest increase in critical thinking would be seen for the students believing to need statistics in the future, as they would engage more in the subject.

The secondary aim is to investigate further the psychometric properties of the brief version of the Motivated Strategies of Learning Questionnaire critical thinking scale (MSLQ; Pintrich et al., 1991) resulting from a recent validation study, which critically considered the content and construct validity of this scale as well as its cross-cultural validity (Nielsen et al., 2021). As this brief critical thinking scale (CTh) was shown to fit the Rasch model both with a Danish and a Spanish sample of psychology students and have reliability for the Danish sample at the level of those obtained with the original scale, we found the CTh scale to be a good candidate for the current study.

MATERIALS AND METHODS

Instrument

The Critical Thinking scale (CTh) employed in the present study is a brief version of the critical thinking scale from the Motivated Strategies of Learning Questionnaire (MSLQ; Pintrich et al., 1991) resulting from a recent validation study, which critically considered the content and construct validity of this scale as well as its cross-cultural validity (Nielsen et al., 2021). The MSLQ is a multi-scale questionnaire intended to measure aspects of students' motivational orientation and learning strategies in high school and higher education (Pintrich et al., 1991). One of the scales included in the MSLQ is a five-item course-specific critical thinking scale with a seven-point response scale anchored for meaning only at the extremes. Of all the short scales in the MSLQ, the critical thinking scale was originally reported as having one of the highest reliabilities (Cronbach's alpha 0.8) with the development sample of 380 Midwestern college students (Pintrich et al., 1991). More recently, Holland et al. (2018) in their meta-analysis found the reliability of the critical thinking scale to be similar across 344 samples ($N = 27,619$) stemming from 32 countries and 14 languages (mean Cronbach's alpha 0.78).

In their study of the cross-cultural validity of the critical thinking scale from the MSLQ, Nielsen et al. (2021) analyzed thoroughly the content validity of the scale and found that only three items actually measured critical thinking.

Content validity was considered both with a theoretically based approach, i.e., analysis of the item content in relation to three different and prevalent definitions of critical thinking (Facione, 1990; Pintrich et al., 1991; Halpern, 2003), and a statistical and psychometric approach, i.e., analysis of local independence and dimensionality (Kreiner and Christensen, 2004). Both approaches reached the conclusion that two items (the same) should be eliminated in order to improve content validity by eliminating construct contamination.

In addition to eliminating two items, Nielsen et al. (2021) also employed an adapted five-point response scale with meaning anchors for all categories with the brief CTh scale in order to pre-assign the meaning that respondents should infer from the categories and thus prevent a random assignment of meaning to a row of numbers, which would affect the validity in interpretation and reliability (Krosnick and Fabrigar, 1997; Maitland, 2009; Menold et al., 2014). This approach was further supported empirically in previous validity studies of other scales from the MSLQ, e.g., Nielsen (2018) with the motivation scales; Nielsen (2020), Nielsen et al. (2017, 2022) with the self-efficacy scale, where a similar adaptation of the response scale had no noteworthy effect on the reliability of the scales compared to the original version.

The three-item CTh scale with the adapted response scale (see below) resulting from the study by Nielsen et al. (2021) had reliability at the level of the original five-item scale with seven response categories for a Danish sample of psychology students (0.82), while slightly lower for a Spanish sample of psychology students (0.73).

The items of the brief CTh scale employed in the present study address the purposeful and inquiring aspect of CTh common to three different definitions of critical thinking (Facione, 1990; Pintrich et al., 1991; Halpern, 2003): how often the student questions things and decide about them (item: I often find myself questioning things I hear or read in this statistics course to decide if I find them convincing); how the student decides about a theory, interpretation or conclusion (item: when a theory, interpretation or conclusion is presented in the statistics course or in the readings, I try to decide if there is good supporting evidence); how the student looks for alternatives (item: whenever I read or hear an assertion or conclusion in this statistics course, I think about possible alternatives) (see also **Supplementary Table A1 in Supplementary Appendix 1**). Thus, the CTh scale does not cover all aspects of critical thinking, but it covers the core aspects, and more importantly, it is not "contaminated" by items not measuring critical thinking (Nielsen et al., 2021). As with the MSLQ, students rate how they feel that the item statements in the brief CTh scale describe them in relation to a specified course (in this case statistics) in terms of frequency of the thinking described in the items: 1 = never, 2 = rarely, 3 = sometimes, 4 = often, and 5 = always. The Danish item texts can be seen in **Supplementary Appendix 1** with the English equivalents (**Supplementary Table A1**). In this article, CTh items are referenced with their original order from the MSLQ to facilitate comparison to other studies with item-level data.

At baseline, students also provided information on gender and age, whether students perceived their mathematical knowledge to be adequate for learning statistics, and whether they believed they would need statistics in their future employment.

Participants and Data Collection

Participants were two consecutive year-cohorts of first-semester students enrolled in a full-time Bachelor of Psychology program in a major Danish university. The students were all taking a two-semester-long statistics course placed in the first two semesters of the bachelor's program. The course consists of weekly lectures and weakly exercise classes. The learning objectives for the first semester of the course contain implicit references to critical thinking (see **Supplementary Appendix 2**). The course has a separate exam in each of the two semesters, and the first-semester exam is an on-campus written exam assessed as pass/fail using a set of specified criteria. As part of these criteria are both implicit and explicit references to critical thinking (see **Supplementary Appendix 2**).

The students completed the CTh scale as part of a larger survey 1 month into their first semester of the course and again 1 month into their second semester of the statistics course. Data were collected in class with a paper-pencil survey. The data collections were arranged with the responsible lecturer before the start of the course. Students were informed ahead of the lecture that the data collection would take place and that it was voluntary to complete the survey. At the point of the data collection, students were informed of the purpose of the overall study, that participation was voluntary, that their data would be treated according to the prevailing data protection regulations, and that they could ask to have their data deleted up to a specified point in time where they would be anonymized. In addition, students were provided with a written information sheet providing the same information as well as contact information for the responsible researcher.

The study sample consisted of 336 students at baseline ($n_{\text{cohort 1}} = 166$, $n_{\text{cohort 2}} = 170$), while the follow-up was completed by 270 students with 165 students who could be matched to their baseline response. The matching rate was determined by circumstances related to student enrollment (drop-out and new enrollment), the matching design (asking students for their student ID in handwriting if they wanted to participate again), and chance (students present in the lecture where data were collected). Thus, as various factors contributed to the missingness of data at follow-up, it could not with any certainty be determined whether data were missing at random or not, though the number of contributing factors makes it more likely that they were missing at random. Likewise, the missingness could not be considered in terms of selection bias, due to the external contributing factors. The mean age of the students at baseline was 22.7 years (SD 4.99) and 81% of the 336 students in the baseline sample identified as female, which is a close match to the official gender distribution of the student admitted to the two particular year-cohorts was 81.3% female students (Ministry of Higher Education and Science, 2021). The gender distribution did not change at follow-up, i.e., 82% of the 165 students in the follow-up sample identified as female.

Statistical Analyses

First, we conducted item analysis using the Rasch measurement model (RM; Rasch, 1960) to establish the psychometric properties of the CTh scale both at baseline and at follow-up. The Rasch model was chosen, as Nielsen et al. (2021) have shown the CTh scale to fit the Rasch model in both a Danish and a Spanish sample. Second, we assessed the changes in CTh scores from the start of the first to the start of the second semester as standardized effect sizes.

Item Analyses by Rasch Models

To investigate the measurement properties of the CTh scale (the secondary issue of the study), item analysis by the Rasch model was conducted first on the baseline sample and subsequently in the follow-up sample to confirm the results. The RM provides optimal measurement properties of scales fitting it (Kreiner, 2007, 2013). These properties include:

1. *Unidimensionality* – the scale measures a single latent construct (Critical Thinking).
2. *Local independence of items* (no LD) – responses to a CTh item depends only on the level of Critical Thinking and not on responses to any of the other items on the scale.
3. *Optimal reliability*, as items are locally independent.
4. *Absence of differential item functioning* (no DIF) – responses to a CTh item depends only on the level of critical thinking and not on persons' membership of subgroups such as gender, age, etc.
5. *Homogeneity* – the rank order of the item parameters/item difficulty is the same for all persons.
6. *Score sufficiency* – the sum score is a sufficient statistic for the person's parameter estimates of Critical Thinking.

Homogeneity and sufficiency are properties only provided by the Rasch model, not any other IRT model. The property of sufficiency is particularly desirable when using the summed raw score of a scale, as it is the usual case with the CTh scale. However, fit to the Rasch model facilitates the use of the person parameter estimates resulting from the measurement model (sometimes termed Rasch-scores), and thus either these or the raw scores can be used in subsequent analysis, as preferred by the individual researcher for their specific purpose.

The overall tests of global homogeneity by comparison of item parameters in low and high scoring groups and overall tests of invariance were conducted as overall tests of fit using Andersen (1973) conditional likelihood ratio test (CLR). The fit of individual items to the Rasch model was tested by comparing the observed item-rest-score correlations with the expected item-rest-score correlations under the RM (Kreiner, 2011). Local independence of items and the assumption of no DIF were tested using Kelderman (1984) conditional likelihood ratio test. DIF was tested in relation to five background variables year cohort (1, 2), gender (female and male), median-split age groups (21 years and younger, 22 years and older), as well as baseline perception of the adequacy of mathematical knowledge to learn statistics (not adequate, adequate), and baseline expectancy to need statistics in future employment (yes, maybe, and no).

Reliability was calculated as Cronbach's alpha (Cronbach, 1951). Targeting (whether items provide information in the area of the scale where the sample population is located) was assessed graphically by item maps as well as numerically by two target indices (Kreiner and Christensen, 2013): the test information target index (the mean test information divided by the maximum test information for theta, and the root mean squared error (RMSE) target index (the minimum standard error of measurement divided by the mean standard error of measurement for theta). Both indices should preferably have a value close to one, as this would indicate the degree to which maximum information and minimum measurement error were obtained, respectively. The target of the observed score and the standard error of measurement (SEM) was also calculated. Items maps are plots of the distribution of the item threshold locations against weighted maximum likelihood estimations of the person parameter locations as well as the person parameters for the population (assuming a normal distribution) and the information function.

Critical values were adjusted for false discovery rate (FDR) arising from conducting multiple statistical tests (i.e., controlling type I errors), whenever appropriate (Benjamini and Hochberg, 1995). As recommended by Cox et al. (1977), we distinguished between weak ($p < 0.05$), moderate ($p < 0.01$), and strong ($p < 0.001$) evidence against the model, rather than applying a deterministic 5% critical limit for p -Values.

Analysis of Differences at Baseline and Analysis of Change

To investigate the primary issue of the study, namely changes in critical thinking, the person parameter estimates resulting from the Rasch models, which have equal distance between any two values, were rescaled to the score range of the instrument and used for baseline differences and in the analysis of change. Differences in mean scores for subgroups of students at baseline were tested within the framework of multiple analyses of variance framework to be able to include grouping variables with more than two categories and test for interaction effects. The change was tested using a paired samples t -test approach and change scores at the group level were calculated as the standardized effect size (ES) (i.e., the difference between baseline and follow-up scores relative to the standard deviation of the baseline scores) (Lakens, 2013; Beauchamp et al., 2015). Subgroups of students were defined by our primary independent variables of interest, i.e., perception of the adequacy of their own mathematical knowledge for learning statistics as well as the students' expectations to need statistics in their future employment. As secondary subgroupings, we included gender and age groups, in order to show whether there were any effects of these on baseline levels of critical thinking or on changes that might be imposed on the primary issues.

Software

The item analyses by Rasch models were conducted using DIGRAM (Kreiner, 2003; Kreiner and Nielsen, 2013), while R was used to produce the item maps. Analyses of variance and

t -tests were conducted using SPSS. Effect sizes were calculated using Excel.

RESULTS

Psychometric Properties at Baseline and Follow-Up

Results of the item analyses (the secondary research issue) showed that the baseline data fitted the Rasch model, and this was also the case with the follow-up data. Thus, there was no evidence against global homogeneity or invariance (Table 1), nor was there any evidence against the fit of the individual items to the Rasch model (Table 2). In addition, we found no evidence against local independence of items (Supplementary Table A2 in Supplementary Appendix 1) and no evidence of differential item functioning relative to year cohort, students' baseline perception of the adequacy of mathematical knowledge to learn statistics, students' baseline expectancy to need statistics in future employment, gender, or age (Supplementary Table A3 in Supplementary Appendix 1). Information on Item thresholds, locations, difficulties, targets, and information is also provided in Supplementary Appendix 1 (Supplementary Table A4).

The targeting of the CTh scale to the student sample was excellent at both baseline and follow-up; slightly better at follow-up with a target information index of 86% at follow-up versus 83% at baseline (Supplementary Table A5 in Supplementary Appendix 1). The level of information is highest where most students are located on the CTh scale at both time points (Supplementary Figure A1 in Supplementary Appendix 1). The reliability of the CTh scale was satisfactory for the purpose of statistical analyses at both baseline and follow-up; 0.72 and 0.75 respectively (Supplementary Table A5 in Supplementary Appendix 1).

The conversion from the summed raw scores of the CTh scale to the estimated person parameters resulting from the Rasch model, as well as these person parameters, estimate rescaled to the original range of the CTh scale are provided in Supplementary Appendix 1 (Supplementary Table A6). This allows users of the scales to choose between using the sum scores, which uses the unit of the scale, or to convert these to any of the person parameters estimates, which are continuous and equidistant scores, as preferred for whatever purpose of use.

Differences in Statistics-Related Critical Thinking at Baseline and Changes in Critical Thinking

The primary research question of the study concerned changes in statistics-related critical thinking from the start of the first semester (baseline) to the start of the second semester (follow-up). As we expected the overall change in critical thinking to differ for subgroups of students dependent on their baseline perception of the adequacy of their own mathematical knowledge for learning statistics as well as their expectation to need statistics in their future employment, we first tested baseline differences. To test whether the expected baseline subgroup

difference in critical thinking could be confirmed, we conducted a multivariate analysis of variance using a backward models search strategy, which included the primary independent variables (i.e., perception of mathematical knowledge as adequate or not and expectation to need statistics in future employment) as well as gender and age and all possible two-way interactions between the independent variables. The results showed that only the two primary independent variables defined significant differences for subgroups of students, and there was no interaction effect. Thus, we present simple tests for differences in critical thinking mean scores for subgroups defined by all four of the background variables in **Table 3**. As expected, students who perceived their mathematical knowledge to be inadequate for learning statistics scored lower on statistics-related critical thinking scores at baseline compared to the students who perceived they had an adequate level of mathematical knowledge ($p < 0.001$). Also as expected, students who did not believe they would need statistics in their future employment scored the lowest on statistics-related critical thinking compared to students who thought they might need or would definitely need statistics in future employment ($p < 0.001$).

We then proceeded to analyze the changes in critical thinking. Absolute individual changes on the CTh scale ranged from -5.3 to 5.1 points on the rescaled logit scale (**Supplementary Table A6** in **Supplementary Appendix 1**), thus showing large individual changes in critical thinking from the first to the second semester (**Figure 1**). The overall standardized effect was small and negative (-0.12), and while there were some variations for student strata defined gender, age, perceived

adequacy of math knowledge to learn statistics, and expectation to need statistics in future employment, effect sizes remained small for all subgroups (**Table 4**). Thus, while there were large absolute changes in the equidistant scores resulting from the Rasch models at the individual level, effect size estimates show that there were only very small and predominantly negative effects. Our expectation that students overall would increase in critical thinking was rejected. The same was the case with our expectation that students, who at baseline did not expect to need statistics in their future employment would increase less in critical thinking than students expecting to need statistics. Only two subgroups of students showed an increase, though small, in critical thinking. These were the male students and students who at baseline perceived their mathematical knowledge as inadequate for learning statistics.

DISCUSSION

The main aim of the study was to explore changes in statistics-related critical thinking from the start of the first semester to the start of the second semester of a two-semester-long statistics course, where the curriculum contains learning objectives implying critical thinking and assessment criteria explicitly requiring critical thinking. The results showed that the student group as a whole has a low mean score of statistics-related critical thinking at baseline (i.e., a mean score of 8.05 within the possible range of 3 to 15) and that there were no significant differences related to gender or age at baseline. In a previous

TABLE 1 | Global tests of homogeneity and invariance for the Critical Thinking Scale at baseline and follow-up.

Tests of fit	Baseline			Follow-up		
	CLR	df	<i>p</i>	CLR	df	<i>p</i>
Global homogeneity ^a	9.0	11	0.622	6.4	11	0.844
Invariance						
Year cohort	16.4	11	0.128	9.8	11	0.553
Math adequacy	10.9	11	0.449	7.4	11	0.768
Stat in Future work	40.8	22	0.009 ⁺	30.2	22	0.113
Gender	14.2	11	0.220	7.3	11	0.775
Age groups	17.4	11	0.097	6.7	11	0.824

CTh, Critical Thinking Scale; CLR, Conditional Likelihood Ratio test.

^aThe test of homogeneity is a test of the hypothesis that item parameters are the same for persons with low or high scores.

⁺The Benjamini-Hochberg adjusted critical level for false discovery rate at the 5% level was $p = 0.0083$ and at the 1% level $p = 0.0017$.

TABLE 2 | Item fit statistics for the Critical Thinking Scale at baseline and follow-up.

Items	Baseline			Follow-up		
	Observed γ	Expected γ	<i>P</i>	Observed γ	Expected γ	<i>p</i>
CTh1	0.47	0.51	0.438	0.61	0.58	0.665
CTh2	0.62	0.52	0.035 ⁺	0.64	0.58	0.336
CTh5	0.47	0.52	0.389	0.51	0.58	0.246

γ = Item-rest score correlations in the form of Goodman and Kruskal's rank correlation for ordinal items.

⁺The Benjamini-Hochberg adjusted critical level for false discovery rate at the 5% level was $p = 0.0111$ and at the 1% level $p = 0.0011$.

TABLE 3 | Mean statistics-related critical thinking scores at baseline.

Group (n)	Mean	SD	p
All students (336)	8.05	1.90	
Gender			
Male (51)	8.36	1.73	
Female (272)	7.97	1.92	0.172
Age groups			
21 years and younger (199)	8.13	1.82	
22 years and older (131)	7.99	2.00	0.500
Math knowledge to learn statistics			
Not adequate (53)	7.01	1.89	
Adequate (281)	8.25	1.85	<0.001
Expect to need statistics in future work life			
Yes (86)	8.49	1.69	
Maybe (196)	8.07	1.81	
No (52)	7.13	2.06	<0.001 ^a

p-Values for “math knowledge to learn statistics” and “expect to need statistics in future work life” are one-sided, due to expectations on the direction of differences.

^aPost hoc pairwise tests showed that it was the group not expecting to need statistics in their future employment that differed significantly from the remaining two groups.

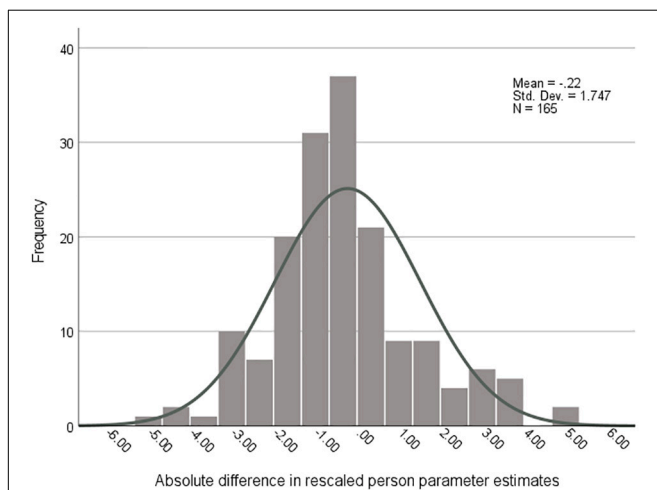


FIGURE 1 | Distribution of differences in Critical Thinking scores (rescaled person parameter estimates) from baseline to follow-up. Differences are shown as follow-up minus baseline so that positive values show an increase and negative values show a decline in critical thinking over time. Distances between any two scores are equal.

cross-cultural study employing the same instrument, statistics-related critical thinking scores were reported at the same level for both Danish and Spanish psychology students, while the mean personality psychology-related critical thinking scores were markedly higher for Danish psychology students, but not the Spanish students (Nielsen et al., 2021). This might very tentatively suggest that domain-specific critical thinking at the start of a semester course varies not only with specific domains with the same academic discipline but also with culture. Two other studies report statistics-related critical thinking at higher levels at the start of a semester course in statistics using different instruments.

Bensley et al. (2010) report medium-level scores on one of their subscales for critical thinking, i.e., the argument analysis scale, at the start of a semester prior to introducing different instructional methods to enhance critical thinking in a research methods course for psychology students. Cheng et al. (2018) report high baseline scores on four single items tapping into four dimensions of critical thinking at the start of introductory statistics classes for students from various academic disciplines. The current results open interesting new avenues of research into domain-specific critical thinking in higher education and its development, both within and between academic disciplines, and across cultures.

Furthermore, we found strong evidence that the baseline statistics-related critical thinking scores differed dependent on students' perception of the adequacy of their mathematical knowledge for learning statistics as well as whether they expected to need statistics in their future work life. Thus, students who perceived their mathematical knowledge to be inadequate for learning statistics had a lower level of critical thinking than students perceiving their mathematical knowledge as adequate. The Danish psychology program requires level B mathematics³ for being admitted to the program but does not require a particular grade for admittance, and thus students can enter with a “just pass”-grade of 02 (see **Supplementary Appendix 2** for the Danish grading scale). As the psychology program is very hard to get into and there is a fixed number of places available, however, only students with a very high-grade point average get in. We assumed that the lack of insight into the field of statistics presents just 1 month into the statistics course and their first semester in the Bachelor of Psychology program might be reflected in their perception of their mathematical basis as adequate or inadequate for learning statistics, and thus also for their inclination toward statistics-related critical thinking at this early point. However, in hindsight, more information on this issue should have been gathered. With regard to baseline differences dependent on the students' expectations to need statistics in their future work life, results were also in line with our assumption, i.e., that confidence in needing statistics in the future would be associated with an enhanced inclination toward statistics-related critical thinking compared to students who were confident they would not need statistics in the future. The results not only confirmed our assumptions but also showed that it was the group of students that were certain to not need statistics in their future work life, who had significantly lowered inclination toward critical thinking compared to both students thinking they might need statistics and students who were sure they would need statistics in the future. The results even showed that there was an ordered relationship in the mean scores for the three groups so that students who expected to need statistics in their future employment had the highest CTh scores, and students who did *not* expect to need statistics in their future employment had the lowest CTh scores and students who thought they might need statistics scored in between. This finding leads us to suggest that future research might explore how the interaction between expectancy-to-need statistics and initial inclination toward statistics-related critical thinking might

³Levels are A, B and C, with A being the highest.

TABLE 4 | Overall and stratified mean differences in critical thinking and effect sizes over time.

Group (n)	Mean difference (ρ)	Effect size	(95% CI)
All students (165)	0.23 (0.050)	-0.12	(-0.33 to 0.10)
Gender			
Male (25)	0.23 (0.313)	0.09	(-0.46 to 0.65)
Female (136)	-0.29 (0.018)	-0.16	(-0.40 to 0.07)
Age groups			
21 years and younger (104)	-0.26 (0.069)	-0.14	(-0.41 to 0.14)
22 years and older (60)	-0.16 (0.233)	-0.08	(-0.44 to 0.27)
Math knowledge to learn statistics			
Not adequate (29)	0.09 (0.405)	0.05	(-0.46 to 0.57)
Adequate (136)	-0.29 (0.021)	-0.15	(-0.39 to 0.09)
Expect to need statistics in future work life			
Yes (45)	-0.46 (0.089)	-0.24	(-0.66 to 0.17)
Maybe (101)	-0.09 (0.283)	-0.05	(-0.32 to 0.23)
No (19)	-0.40 (0.145)	-0.22	(-0.85 to 0.42)

Mean differences are shown as follow-up minus baseline so that positive values show an increase and negative values show a decline in critical thinking over time. *P*-values are one-sided. CI, Confidence Interval. Effect sizes are calculated using the rescaled person parameter estimates, as the distance between any two scores is equal.

be related to the outcome of statistics courses, but also to the actual need for statistics in the first employment of the graduates.

Turning to the main results of the study, namely the lack of an overall increase in statistics-related critical thinking in the first semester, this was the opposite of what we expected. Previous research on the development of statistics-related critical thinking has mainly focused on comparing teaching methods designed to enhance critical thinking with “usual” teaching methods not designed for this purpose, or by simply evaluating the enhancing effect of purposely designed teaching methods. While methods for measuring statistics-related critical thinking differ across studies as does the teaching methods evaluated results are also ambiguous, as some find no effect of the purposely designed teaching compared to the usual teaching without clarifying whether this means there was an effect or no effect for both groups (Goode et al., 2018; Setambah et al., 2019), and others a positive effect for only the students receiving the purposely designed teaching and no change for the students receiving the usual teaching (e.g., Bensley et al., 2010). On the same note, one study evaluating just the effect of a purposely designed teaching method in itself found this to enhance the statistics-related critical thinking of the students (Cheng et al., 2018). The lack of increase in the statistics-related critical thinking in the current study is thus only supported by Bensley et al. (2010), who did not find any change for their control group of psychology students. The current study is not enough to refute that meaningful instruction within a subject domain inherently will entail the development of critical thinking skills even if these are not purposely targeted with teaching activities, as suggested by Tiruneh et al. (2017). However, the current study does show that even the students have a low level of critical thinking at baseline and thus ample room for improvement, one semester’s worth of university-level teaching in statistics with lectures as well as small exercise classes, where assessment criteria explicitly mention critical thinking (**Supplementary Appendix 2**), does not enhance the

critical thinking of the students as a whole. Thus, Tiruneh et al.’s (2017) notion cannot be supported by our research, as we do not find an overall positive effect on statistics-related critical thinking over the semester. Our study, however, points to the need for developing further research to explore the factors involved in the development of statistics-related critical thinking skills.

The subgroup results in the current study also showed small effects for all subgroups, and thus did not divulge any clear patterns with regard to student factors related to the development of statistics-related critical thinking. The findings, which might suggest areas of interest for future research are the differences in the direction of the development in statistics-related critical thinking found across gender and across perceptions of the adequacy of mathematical knowledge for learning statistics, even if these differences in direction might be random results due to small group sizes. Thus, future research should include additional student characteristics to explore this further, e.g., characteristics such as dispositional characteristics such as personality, e.g., conscientiousness which has consistently been found to be positively associated with academic success in higher education (Richardson et al., 2012; Vedel, 2014), an association, which in relation to learning statistics might very well be mediated by statistics-related critical thinking. Motivation and academic self-efficacy, as both have been linked to student performance (Richardson et al., 2012) and student anxiety (Tahmassian and Jalali Moghadam, 2011; Nguyen and Deci, 2016) and statistics-related anxiety is well-documented among students from other disciplines taking statistics courses and the detrimental effect of anxiety on learning is well-known. We thus propose that motivational factors as well as the belief in one’s own ability to learn statistics might moderate the development of statistics-related critical thinking and that this is certainly worth investigating in the future.

Dispositional measures and other student characteristics might also be successfully employed in future studies of increases

and decreases at the individual level, and preferably with more points of measurement (three to six), as they might then contribute to explaining individual student trajectories with regard to statistics-related critical thinking and whether these are one-directional across multiple points of measurement. Such student characteristics might also be useful with larger samples to explore whether certain student profiles are associated with an increase and certain profiles with a decrease in statistics-related critical thinking. In addition, future studies might link to the current research and expand these by including students from other academic disciplines than psychology.

The study has four major strengths. The first strength is that the results concerning change stand on a very strong psychometric foundation as the CTh scale fitted the Rasch model both at baseline and at follow-up and as the scale was very well targeted to the study population of first-year Danish Psychology Bachelor students taking their statistics course. As such, we know that the CTh scale possesses the psychometric properties, we aimed for and that the results of the change analyses and both the differences at baseline and the effect sizes are not biased due to a general lack of invariance or differential item function. The second strength lies in the use of standardized effect sizes to assess changes in statistics-related critical thinking, as this makes it possible for future studies using the same instrument under different conditions to compare the results. The third strength of the study is its contribution to the body of knowledge on the so-called “natural” development of domain-specific critical thinking, by showing that there was no overall increase in critical thinking. The contribution is important, as it showed that even though critical thinking was explicitly mentioned in the assessment criteria and implicitly in the learning objectives for the course as well as the overall competencies to be achieved through the program, no overall increase was found nor were there subgroup-specific increases of any significance. However, equally important is the finding that there were rather large absolute changes in critical thinking at the individual level, both in the form of increases and decreases, as are the findings of baseline differences dependent on students’ perception of the adequacy of their mathematical knowledge for learning statistics and their expectancy to need statistics in their future work life.

Likewise, the study has three limitations. The first is the sample size and the subgroup distributions in the longitudinal sample, as this did not allow us to explore any possible interaction effects by stratifying on more than one grouping variable at a time. Thus, it was not possible to explore with any certainty how the differences in statistics-related critical thinking at baseline might affect the development. The second limitation might be considered to be the CTh scale itself, as it only comprises three items covering the purposeful and inquiring aspect of CTh common to three major definitions of critical thinking. However, as thoroughly demonstrated with the content and construct validity analyses conducted by Nielsen et al. (2021), there is no loss in content validity by eliminating two items from the original scale from the MLSQ, as these did in fact not measure critical thinking – not content-wise nor when considering the dimensionality issue. As the brief version, we employed in this study, furthermore fitted the strictest measurement model (i.e., the Rasch model) and was

well targeted to the student population in this study and the cross-cultural sample in the study by Nielsen et al. (2021), we do not find the brief version to be inferior to the five-item version from the MLSQ, rather the contrary. However, we do recognize that other and longer instruments might be preferred by other researchers and that such instruments, if appropriately validated, can offer more precise measurement. The third limitation is that we did not collect any qualitative and detailed information from the professor or the students, which might have contributed to a better understanding of the lack of overall increase in statistics-related critical thinking as well as the results at the individual level.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: [10.5281/zenodo.6401225](https://doi.org/10.5281/zenodo.6401225).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

TN construed the study and conducted the analyses. All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

The authors have received to funding for the study.

ACKNOWLEDGMENTS

The authors thank the students and the professor who kindly lend their time to provide data, Martin Andersen for collecting part of the data, Pedro Henrique Ribeiro Santiago for providing the R-code for the item maps, and Mary McGovern for translating the part of the curricula descriptions which was only available in Danish.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/educ.2022.884635/full#supplementary-material>

REFERENCES

- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika* 38, 123–140. doi: 10.1007/BF02291180
- Bailin, S., Case, R., Coombs, J. R., and Daniels, L. B. (1999). Common misconceptions of critical thinking. *J. Curric. Stud.* 31, 269–283. doi: 10.1080/002202799183124
- Beauchamp, M. K., Jette, A. M., Ward, R. E., Kurlinski, L. A., Kiely, D., Latham, N. K., et al. (2015). Predictive validity and responsiveness of patient-reported and performance-based measures of function in the Boston RISE study. *J. Gerontol. Med. Sci.* 70, 616–622. doi: 10.1093/gerona/glu227
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc. Series B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bensley, D. A., and Baxter, C. (2006). *The Critical Thinking in Psychology Test. Unpublished manuscript.* Frostburg, MD: Frostburg State University.
- Bensley, D. A., Crowe, D. S., Bernhardt, P., Buckner, C., and Allman, A. L. (2010). Teaching and Assessing Critical Thinking Skills for Argument Analysis in Psychology. *Teach. Psychol.* 37, 91–96. doi: 10.1080/00986281003626656
- Cheng, S., Ferris, M., and Perolio, J. (2018). An innovative classroom approach for developing critical thinkers in the introductory statistics course. *Am. Statist.* 72, 354–358. doi: 10.1080/00031305.2017.1305293
- Cox, D. R., Spjøtvoll, E., Johansen, S., van Zwet, W. R., Bithell, J. F., and Barndorff-Nielsen, O. (1977). The Role of Significance Tests [with Discussion and Reply]. *Scand. J. Stat.* 4, 49–70.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1016/0020-7489(93)90092-9
- De Jager, T. (2012). Can first year students' critical thinking skills develop in a space of three months?. *Procedia. Soc. Behav. Sci.* 47, 1374–1381. doi: 10.1016/j.sbspro.2012.06.829
- Facione, P. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations.* Newark, NJ: American Philosophical Association.
- Facione, P. A., Facione, N. C., and Giancarlo, C. A. F. (1998). *The California Critical Thinking Disposition Inventory.* California: Academic Press, 67–79.
- Goode, C. T., Lamoreaux, M., Atchison, K. J., Jeffress, E. C., Lynch, H. L., and Sheehan, E. (2018). Quantitative Skills, Critical Thinking, and Writing Mechanics in Blended Versus Face-to-Face Versions of a Research Methods and Statistics Course. *Teach. Psychol.* 45, 124–131. doi: 10.1177/0098628318762873
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *Am. Psychol.* 53, 449–455. doi: 10.1037/0003-066X.53.4.449
- Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking.* Mahwah, NJ: Erlbaum.
- Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking. 5th Edn.* New York, NY: Psychology Press Taylor & Francis Group.
- Halpern, D. F. (2015). *Halpern critical thinking assessment.* Austria: Schuhfried GmbH.
- Hammer, S. J., and Green, W. (2011). Critical thinking in a first year management unit: the relationship between disciplinary learning, academic literacy and learning progression. *Higher Educ. Res. Dev.* 30, 303–315. doi: 10.1080/07294360.2010.501075
- Holland, D. F., Kraha, A., Zientek, L. R., Nimon, K., Fulmore, J. A., Johnson, U. Y., et al. (2018). Reliability Generalization of the Motivated Strategies for Learning Questionnaire: a Meta-Analytic View of Reliability Estimates. *SAGE Open* 8, 1–29. doi: 10.1177/2158244018802334
- Kanbay, Y., Isik, E., Aslan, O., Tektas, P., and Kilic, N. (2017). Critical Thinking Skill and Academic Achievement Development in Nursing Students: four-year Longitudinal Study. *Am. J. Educ. Res. Rev.* 2;12. doi: 10.28933/ajerr-2017-12-0501
- Kaya, H., Şenyuva, E., and Bodur, G. (2017). Developing critical thinking disposition and emotional intelligence of nursing students: a longitudinal research. *Nurse Educ. Today* 48, 72–77. doi: 10.1016/j.nedt.2016.09.011
- Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika* 49, 223–245.
- Kreiner, S. (2003). *Introduction to DIGRAM.* Copenhagen: Department of Biostatistics, University of Copenhagen.
- Kreiner, S. (2007). Validity and objectivity. Reflections on the role and nature of Rasch Models. *Nordic Psychol.* 59, 268–298. doi: 10.1027/1901-2276.59.3.268
- Kreiner, S. (2011). A Note on Item-Restscore Association in Rasch Models. *Appl. Psychol. Meas.* 35, 557–561. doi: 10.1177/014662161141022
- Kreiner, S. (2013). “The Rasch model for dichotomous items,” in *Rasch Models in Health*, eds K. B. Christensen, S. Kreiner, and M. Mesbah (London: ISTE Ltd, Wiley), 5–26. doi: 10.1002/9781118574454.ch1
- Kreiner, S., and Christensen, K. B. (2004). Analysis of local dependence and multidimensionality in graphical loglinear Rasch models. *Commun. Stat. Theory Methods* 33, 1239–1276. doi: 10.1081/sta-120030148
- Kreiner, S., and Nielsen, T. (2013). *Item analysis in DIGRAM 3.04. Part I: Guided tours. Research report 2013/06.* Denmark: University of Copenhagen, Department of Public Health.
- Kreiner, S. and Christensen, K. B. (2013). “Person Parameter Estimation and Measurement in Rasch Models”, in *Rasch Models Health*, eds K.B. Christensen, S. Kreiner and M. Mesbah. (London, ISTE and John Wiley & Sons, Inc.) 63-78. doi: 10.1002/9781118574454.ch4
- Krosnick, J. A., and Fabrigar, L. R. (1997). “Designing rating scales for effective measurement in surveys,” in *Survey measurement and process quality*, eds L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, et al. (New York, NY: John Wiley), 141–164. doi: 10.1002/9781118490013.ch6
- Kuhn, D. (1999). A developmental model of critical thinking. *Educ. Res.* 28, 16–25. doi: 10.2307/1177186
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863
- Lau, J. Y. F. (2015). “Metacognitive education: Going beyond critical thinking,” in *The palgrave handbook of critical thinking in higher education*, eds M. Davies and R. Barnett (New York, NY: Palgrave Macmillan), 373–389. doi: 10.1057/9781137378057
- Maitland, A. (2009). Should I label all scale points or just the end points for attitudinal questions? *Survey Pract.* 4, 1–4. doi: 10.29115/SP-2009-0014
- McGuirk, J. (2021). Embedded rationality and the contextualization of critical thinking. *J. Philosop. Educ.* 55, 606–620. doi: 10.1111/1467-9752.12563
- McPeck, J. (1992). “Thoughts on subject specificity,” in *The generalizability of critical thinking: Multiple perspectives on an educational ideal*, ed. S. Norris (New York, NY: Teachers College Press), 198–205.
- Menold, N., Kaczmarek, L., Lenzen, T., and Neusar, A. (2014). How Do Respondents Attend to Verbal Labels in Rating Scales? *Field Methods* 26, 21–39. doi: 10.1177/1525822X13508270
- Ministry of Higher Education and Science (2021). *Ansøgere og optagne fordelt på køn, alder og adgangsgrundlag.* Available online at: <https://ufm.dk/uddannelse/statistik-og-analyser/sogning-og-optag-pa-videregaende-uddannelser/grundtal-om-sogning-og-optag/ansogere-og-optagne-fordelt-pa-kon-alder-og-adgangsgrundlag> (accessed date 23.12.2021).
- Moore, T. (2011). Critical thinking and disciplinary thinking: a continuing debate. *High. Educ. Res. Dev.* 30, 261–274. doi: 10.1080/07294360.2010.501328
- Moseley, D., Baumfield, V., Elliott, J., Higgins, S., Miller, J., Newton, D. P., et al. (2005). *Frameworks for thinking: A handbook for teaching and learning.* Cambridge, UK: Cambridge University Press.
- Nguyen, T. T., and Deci, E. L. (2016). Can it be good to set the bar high? The role of motivational regulation in moderating the link from high standards to academic well-being. *Learn. Individ. Diff.* 45, 245–251. doi: 10.1016/j.lindif.2015.12.020
- Nielsen, T. (2018). The intrinsic and extrinsic motivation subscales of the Motivated Strategies for Learning Questionnaire: a Rasch-based construct validity study. *Cog. Educ.* 5, 1–19. doi: 10.1080/2331186X.2018.1504485
- Nielsen, T. (2020). The Specific Academic Learning Self-efficacy and the Specific Academic Exam Self-Efficacy scales: construct and criterion validity revisited using Rasch models. *Cog. Educ.* 7, 1–15. doi: 10.1080/2331186X.2020.1840009
- Nielsen, T., Makransky, G., Vang, M. L., and Dammeyer, J. (2017). How specific is specific self-efficacy? A construct validity study using Rasch measurement models. *Stud. Educ. Eval.* 57, 87–97. doi: 10.1016/j.stueduc.2017.04.003
- Nielsen, T., Martínez-García, I., and Alastor, E. (2021). Critical Thinking of Psychology Students: a Within- and Cross-Cultural Study using Rasch models. *Scand. J. Psychol.* 62, 426–435. doi: 10.1111/sjop.12714
- Nielsen, T., Martínez-García, I., and Alastor, E. (2022). “Psychometric properties of the Spanish translation of the Specific Academic Learning Self-Efficacy and the Specific Academic Exam Self-Efficacy scales in a higher education context,” in *Academic Self-efficacy in Education: Nature, Measurement, and Research*, eds

- M. S. Khine and T. Nielsen (New York, NY: Springer), 71–96. doi: 10.1007/978-981-16-8240-7_5
- Özelçi, S. Y., and Çalışkan, G. (2019). What is critical thinking? A longitudinal study with teacher candidates. *Internat. J. Eval. Res. Educ.* 8, 495–509. doi: 10.11591/ijere.v8i3.20254
- Paul, R., and Elder, L. (2005). *A guide for educators to Critical Thinking Competency Standards*. Santa Barbara, CA: Foundation for Critical Thinking.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., and McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. (Technical Report No. 91-8-004). Ann Arbor, MI: The Regents of the University of Michigan.
- Ralston, P. A., and Bays, C. L. (2015). Critical thinking development in undergraduate engineering students from Freshman Through Senior Year: a 3-Cohort Longitudinal Study. *Am. J. Eng. Educ.* 6, 85–98. doi: 10.19030/ajee.v6i2.9504
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Ren, X., Tong, Y., Peng, P., and Wang, T. (2020). Critical thinking predicts academic performance beyond general cognitive ability: evidence from adults and children. *Intelligence* 82:101487. doi: 10.1016/j.intell.2020.101487
- Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol. Bull.* 138, 353–387. doi: 10.1037/a0026838
- Ricketts, J. C., and Rudd, R. D. (2005). Critical thinking of selected youth leaders: the efficacy of critical thinking dispositions, leadership and academic performance. *J. Agricult. Educ.* 46, 32–43.
- Saenab, S., Zubaidah, S., Mahanal, S., and Lestari, S. R. (2021). ReCODE to Re-Code: an instructional model to accelerate students' critical thinking skills. *Educ. Sci.* 11:2. doi: 10.3390/educsci11010002
- Sahanowas, S. K., and Halder, S. (2020). Critical thinking disposition of undergraduate students in relation to emotional intelligence: gender as moderator. *Heliyon* 6:e05477. doi: 10.1016/j.heliyon.2020.e05477
- Setambah, M. A. B., Tajudin, N. M., Yaakob, M. F. M., and Saad, M. I. M. (2019). Adventure Learning in Basics Statistics: impact on Students Critical Thinking. *Internat. J. Instruct.* 12, 151–166. doi: 10.29333/iji.2019.12310a
- Stupnisky, R. H., Renaud, R. D., Daniels, L. M., Haynes, T. L., and Perry, R. P. (2008). The interrelation of first-year college students' critical thinking disposition, perceived academic control and academic achievement. *Res. High. Educ.* 49, 513–530. doi: 10.1007/s11162-008-9093-8
- Tahmassian, K., and Jalali Moghadam, N. (2011). Relationship between self-efficacy and symptoms of anxiety, depression, worry and social avoidance in a normal sample of students. *Iran. J. Psychiatry Behav. Sci.* 5, 91–98.
- Thomas, T. (2011). Developing first year students' critical thinking skills. *Asian Soc. Sci.* 7, 26–35. doi: 10.5539/ass.v7n4p26
- Tiruneh, D. T., De Cock, M., and Elen, J. (2017). Designing Learning Environments for Critical Thinking: examining Effective Instructional Approaches. *Internat. J. Sci. Mathem. Educ.* 16, 1065–1089. doi: 10.1007/s10763-017-9829-z
- Vedel, A. (2014). The Big Five and tertiary academic performance: a systematic review and metaanalysis. *Personal. Individ. Diff.* 71, 66–76. doi: 10.1016/j.paid.2014.07.011
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Nielsen, Martínez-García and Alastor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.