Educational Effectiveness: Validation of an Instrument to Measure Students' Critical Thinking and Disposition

Tina Gerdts-Andresen
Østfold University Collage, Norway, *tina.gerdts-andresen@hiof.no*

Mette Tindvik Hansen
Østfold University Collage, Norway, *mette.t.hansen@hiof.no*

Vigdis Abrahamsen Grøndahl
Prof., Østfold University Collage, Norway, *vigdis.a.grondahl@hiof.no*

Critical thinking has become an educational ideal in a higher education context, where the goal is to develop students' critical and reflective attitudes. There is a need for validated tools to raise awareness of the student's perception of their critical thinking, and to measure whether teaching has a positive impact on students' ability to think critically and to increase the students' disposition towards critical thinking. The aim of this study was to adapt and validate the Critical Thinking Disposition Scale (CTDS) for use in social work education in Norway, and to explore students' perceptions of their critical thinking. CFA was conducted for psychometric evaluation, and Cronbach's alpha was used to evaluate internal consistency. Students' perception of their critical thinking was analysed at item and factor level. CTDS-NO show satisfying psychometric properties with high reliability and validity and may be a valuable tool for measuring students' critical thinking disposition during their education. Results indicate that the questionnaire may also be a useful tool for measuring whether the teaching contributes to the development of critical thinking students.

Keywords: critical thinking, students' perception, dispositions, psychometric validation, scale translation, confirmatory factor analysis

INTRODUCTION

This study introduces validation of an instrument to awake awareness on and measure student's critical thinking and disposition. The idea of critical thinking stems from an educational tradition, most often associated with John Dewey (Boot & Reynolds, 1983). This is attributed to the theory that learning occurs through experience, which involves exploring one's understanding of one's own actions. However, in recent times critical thinking has most often been associated with Schön (1983), who developed a model for reflective practice and by this forwarded the idea that critical thinking about one's own practice may lead to better practice. With this, critical thinking has become an

educational ideal in a teaching context where the goal is to develop students' critical and reflective attitudes through the use of critical thinking (Ennis, 2008; Sosu, 2013).

There are different ways of defining 'critical thinking' today, but the most commonly used is the recognition that thought processes are composed of cognitive and dispositional dimensions (Nugroho et al., 2018). Critical thinking includes seeing the different perspectives of a relationship and openness towards other perspectives, as well as the support of thoughts and feelings concerning actual relationships (Brookfield, 2012; Hooks, 2010; Merriam & Bierema, 2014). According to Brookfield (2012), there are three aspects to critical thinking: 'critical theory', 'critical thinking' and 'critical action'. Critical theory is a critique of social conditions, whereas critical thinking is a philosophy that focuses on how hegemony or dominant groups create 'truths' that are accepted as natural and right. This has influenced critical perspectives such as postmodernism, feminist pedagogy, critical race theory, and multiculturalism (Brookfield, 2012; Merriam & Bierema, 2014).

Critical thinking is a social learning process (Brookfield, 2014: 55) in which students are trained to take a critical perspective of themselves through their own filters. It is about reflecting on one's own values and assumed truths, set against one's own actions. The goal is to create self-realised professionals (Hooks, 2010). According to Brookfield (2012), such learning is best promoted when teaching topics linked to specific events (experiences through case studies) and when students are dealing with an unexpected event that challenges their understanding of learned knowledge.

The third aspect, critical action, is the ability to take thoughtful action when you have critically evaluated your thoughts, skills and opportunities. In teaching, this is about helping students relate to critical theory, first individually (through teaching), and then by connecting their own experience with that of others' and making a qualified choice of action in a specific situation (Brookfield, 2012). It is about reflecting on an experience or planned action and any possible outcomes in the future and then evaluating the experience and considering what could be done differently. Dewey (Boot & Reynolds, 1983) calls the relationship between thinking and experience 'the method of intelligent learning'. He refers here to an understanding that no experience has meaning without elements of thinking and that critical thinking is a central component of experiential learning. Good command of critical thinking is the presence of analytical, interpretive, explanatory, self-regulating, and evaluative activities (Rahdar et al., 2018; Suryanti & Nurhuda, 2021). It is about training the students' ability to acquire new knowledge produced and reproduced by themselves (Rahdar et al., 2018).

Measuring Students' Critical Thinking

Student learning outcomes have been traditionally measured using a test developed to measure generic knowledge, i.e. critical thinking (Melguizo et al., 2017). In line with the expanding focus on developing students' ability to think critically, various strategies and learning methods are used to create this capability (Suryanti & Nurhuda, 2021). However, there is a need for validated tools to measure whether the strategies and

methods of learning have the intended positive impact and increases students' awareness of critical thinking.

Studies show that students are quite confident in their ability to think critically, but with a rather narrow view of the skills and dispositions that constitute this (Orszag, 2015). According to Poondej and Lerdpornkulrat (2015), there is no agreement on how to best measure the students' disposition towards critical thinking, and not many instruments are designed for measuring both critical thinking and students' disposition. Database searches have identified a handful of validated questionnaires, but none of the identified questionnaires are translated into the Norwegian language or context. Moreover, very few of these questionnaires measure both cognitive and dispositional dimensions. Throughout the literature, there has been a varying concept of components for critical thinking disposition. However, philosophers, psychologists and educators tend to focus on similar sets, and the most common components of these dispositions are open-mindedness, an interest in other points of view, curiosity, the disposition to seek and evaluate reasons, and objectivity (Poondej & Lerdpornkulrat, 2015).

A recently developed questionnaire by Sosu (2013) is the 11 question 'Critical Thinking Disposition Scale' (CTDS). Items for the scale were generated by Sosu (2013) using taxonomies of important thinking dispositions discussed in various literature. To ensure a comprehensive coverage of key dispositions, Sosu (2013) generated an initial pool of 98 items based on the identified taxonomies. Where there was a lack of consensus within the literature that a particular disposition constituted 'critical thinking', the item was excluded. After eliminating duplicates, 46 items remained. These 46 items were structured as five-point Likert-scale responses for piloting. Through the piloting process, several items were reworded, flagged, or deleted, leading to 24 items remaining for empirical testing. During the statistical analysis of the empirical testing, items that appeared as single indicators, without correlation to other items, was assessed and eliminated. Thereafter, each item was undertaken a critical review to ensure a wording free from ambiguity, leading to elimination of 11 items. An additional two items were thereafter deleted as they cross-loaded and showed values below a defined cut-off value of .30 (Sosu, 2013). The result was an 11 item questionnaire. While the questionnaire validation study supported the two-factor dimension of a United Kingdom respondent sample, Sosu (2013) argues that 'there is a need for further validation work that includes the convergent validity of the scale as well as a broader study of whether a two-factor dimension demonstrates concurrent validity for respondents from other countries'. This study contributes to the literature by including both the convergent validity of the CTDS (Sosu, 2013) scale as well as a broad approach to the factor dimension.

Objective

The aim of this study is two-fold: First, to adapt and validate the CTDS for use in social work education in Norway. Second, to provide evidence for its reliability and validity as a tool to measure the educational effectiveness of students' critical thinking disposition by exploring students' perceptions of their critical thinking.

METHOD

Study Design

This is an adaption and validation study with a qualitative cross-sectional design.

Setting and Respondents

The participants are Norwegian undergraduate students majoring in social work. Out of approximately 262 students, 217 were invited to participate in the validation process of the Norwegian translated version of CTDS during two lectures, one in February 2020 and one in November 2019. Number of students that consented to participation by filling out the questionnaire was 197. Only questionnaires that were fully completed were included in the statistical analysis. The total number of participants included in this study for further analysis was 182. According to Tabachnick and Fidell (2014), the sample size needs to be more than five times the number of items in the questionnaire in order to be representative of the adapted questionnaire's validation. With a questionnaire of 11 items, one would require a minimum of 55 participants, corresponding to 30% of the sample size used in this study.

Measures

The CTDS (Sosu, 2013) is originally an English instrument developed to measure students' disposition to critical thinking. The participants are asked to take a stand on 11 statements that deal with how they reflect and resonate. The 11 statements are thereafter grouped into two domains: (1) 'Critical Openness' (7 items) and (2) 'Reflective Scepticism' (4 items). A 'five-point Likert-type response' scale was used with '1' ('strongly disagree') and '5' ('strongly agree') as endpoints. The original format and procedures in giving instructions and presenting test items is equally familiar to both populations in England and Norway. The Norwegian translation of the questionnaire and its instructions was therefore performed close to its original form (CTDS), as recommended by Hambleton (2005) when adapting educational and psychological tests for cross-cultural assessment. The general instruction from CTDS (Sosu, 2013) was adapted, emphasising that the questions covered neither the actual behaviour nor the desirability of the behaviour. The participants are asked to take a stand on 11 statements that deal with how they reflect and resonate.

Procedures and Data Collection

The goal of the adaption process was to achieve construct equivalence between the original questionnaire (CTDS; Sosu, 2013) and the translated Norwegian version (CTDS-NO) by adapting both test instructions and items. Construct equivalence includes both correlation in the way the construct measured by the test is operationalised in each language and cultural group and conceptual/functional equivalence (Harkness, 1998). The adaption process is a combination of close translation of those parts of the instrument that are believed to be sufficient in the target culture, and a change of other parts when a closer translation would be inadequate for cultural, linguistic, or psychometric reasons (Hambleton & De Jong, 2003; Hambleton, 2005; Harkness,

Mohler & Van de Vijver, 2003). The adaption process were performed in accordance with the International Test Commission (ITC) Guidelines (Hambleton, 2005).

To achieve substantive similarity, the original questionnaire (CTDS) was translated from its original language ('English') to Norwegian according to the procedure 'Guidelines for the Process of Cross-Cultural Adaptation of Self-Reports' (Guillemin et al., 1993). Initially, two translations were carried out, back and forth between English and Norwegian, where each pair of translators had bilingual competence, consisting of a native English speaker and a native Norwegian speaker. As recommended by Hambleton (2005) and Sireci (2005), both translator pairs were experienced in test development, but also familiar with the content area. The translators' evaluated the differences between the translated versions and achieved satisfactory compliance with the original scale. Two academics in the research field evaluated the completed Norwegian version for cultural appropriateness. The updated version was then re-evaluated and assessed in relation to the original version. When designing the final version, the content of the original questionnaire was weighted more heavily than the direct meaning of the words that were translated, as recommended when adapting tests for cross-cultural assessment (Hambleton & De Jong, 2003; Hambleton, 2005; Polit and Beck, 2016).

After the translation, a pilot study was carried out on a group of 98 Norwegian second year undergraduate social work students to test the CTDS-NO for bias. In cross-cultural research, bias is a generic term for all kinds of factors that threaten intergroup comparisons' validity (Van de Vijver & Hambleton, 1996). CTDS-NO was tested for three types of bias: construct bias, method bias, and item bias. Construct bias refers to differences in constructs across cultural groups (e.g. behaviours, attitudes, norms). Method bias is a generic term that includes all nuisance variables due to method-related factors, such as all score differences and instrument properties that arise due to difficulties with the instruction. Item bias, or differential item functioning, refers to validity threats that affect separate items in the questionnaire. The most common item bias source is a poor translation and different connotations (Van De Vijver & Poortinga, 2005).

After completing the questionnaire, the students attended a workshop where they gave their opinion on the relevance, acceptability and understandability of the instructions and items. The pilot study identified that both instructions and the items were comprehensible and that the items comprised the students' critical thinking disposition. However, there was some reflection on the choice of words, which lead to some minor adjustments. There was also a question regarding the suitability of items 6 and 7. Item 6 draws attention to the exploration of another person instead one's own opinions or position, while item 7 focuses on the defence of action rather than reflection on disposition. Both items remained in the questionnaire for further assessment.

The CTDS-NO was then administered to students from social work bachelor programmes in Norway. The students were invited to participate by one of the academics in the research field during a lecture. The students were provided with both oral and written information regarding their participation, explaining that participation was voluntary and all answers would be treated confidentially. The students' gave their

consent by filling out the questionnaire and by that participated in the study. All participants filled out the questionnaire on paper and in person. See the Appendix for the Norwegian version of the CTDS-NO.

Statistical Analysis

Descriptive data were analysed using IBM SPSS version 25 (Chicago, Illinois, USA). Data were checked for normality both by assessing skewness and kurtosis and graphically. The number of missing values was less than one percent. Paired sample t-tests were used for comparisons of the dimensions in the CTDS-NO instrument, and p-values below .05 were defined as statistically significant. Cronbach's alpha was used to establish internal consistency, with values above .70 indicating an acceptable level of reliability (Cronbach, 1951; Streiner & Norman, 2008).

When assessing construct equivalence across assessment instruments, there are at least four statistical approaches (Sireci, Patsula & Hambleton, 2005). Confirmatory Factor Analysis (CFA) is one such approach, and the statistical approach that is the most commonly used when testing the overall goodness-of-fit of a specific factor dimension or to verify number of underlying dimensions of an instrument and the pattern of item-factor relationships (factor loadings; Brown, 2015). To determine the factorability of the data, 'Bartlett's test of sphericity' and the 'Kaiser-Meyer-Olkin (KMO)' was performed (Field, 2017; Hair et al., 2014). For factor analysis to be considered appropriate, 'Bartlett's test of sphericity' should be significant (p <.05). KMO measures sampling adequacy with .6 suggested as the minimum value for factor analysis (Tabachnick & Fidell, 2014). Based on the theoretical model underlying the CTDS instrument developed for students enrolled in a programme within education, the extraction was guided through two factors with oblique rotation produced by the Oblimin with Kaiser Normalisation (Watkins, 2006).

The construct validity of a two-dimensional model was tested by running a CFA by using AMOS V.25 for SPSS. The CFA was interpreted using 'goodness-of-fit indices'. Issues surrounding 'goodness-of-fit indices' are hotly debated, as fit indices are often differentially affected by various aspects of the analytic situation, such as sample size, model complexity, estimation method, amount and type of misspecification, normality of data, and type of data. Thus, based on Hu and Bentler (1999) and Brown (2015), there are a few consensus areas regarding recommended fit index cutoff criteria and 'goodness-of-fit' guidelines. The classic 'goodness-of-fit' index is χ^2/df (chi-square/degree of freedom). It is, however, rarely used in applied research as a sole index of model fit. While χ^2/df is routinely reported in CFA research, other fit indices are usually relied on more heavily when evaluating model fit. Following the fit indices, presented in the original CTDS instrument analysis (Sosu, 2013), support of contentions of a reasonably good fit between CTDS and CTDS-NO is obtained in instances where: (1) The standardised root mean square residual (SRMR) values are close to .08 or below; (2) The root mean square error of approximation (RMSEA) values are close to .06 or below; (3) Comparative fit index (CFI), and Tucker-Lewis index (TLI) values are .90 or above. Values of Chi-Square test > .05 do not represent any significant

association between the subscales and indicate a good fit factor dimension (Hair et al, 2014).

Ethics Approval and Consent to Participate

All data were collected anonymously. In Norway, studies that involve exclusively questionnaire-based data are not subjective to ethical approval by the data protection services (NSD). They were, however, contacted and confirmed that the study did not require reporting.

FINDINGS

Descriptive statistics for the CTDS-NO at item level are presented in Table 1. The social work students' perceptions of their critical thinking disposition had a mean level range from 3.32 – 4.52 on the subscale *'critical openness',* and 3.73 – 4.00 on the subscale *'reflective scepticism'*, with low measures of dispersion on both subscales. There is a ceiling effect and a high level (> .07) of skewness and kurtosis in items 3 and 6. The ceiling effect indicates that the students have high belief in their critical thinking disposition, both regarding their '*critical openness*' and with respect to their '*reflective scepticism*'.

Table 1
Descriptive statistics at the item level (N=182)

Factors	Items	Mean	SD
Factor 1: Critical Openness	'N'* (1) 'I usually try to think about the bigger picture during av discussion'	3,47	.66
	'C'* (2) 'I often use new ideas to shape (modify) the way I do things'	3,32	.77
	'D'* (3) 'I use more than one source to find out information for myself'	3,92	.88
	'A'* (4) 'I am often on the lookout for new ideas'	3,64	.85
	'O'* (5) 'I sometimes find a good argument that challenges some of my firmly held beliefs'	3,61	.74
	'J'* (6) 'It's important to understand other people's viewpoint on an issue'	4,52	.60
	'E'* (7) 'It is important to justify the choices I make'	3,97	.77
Factor 2: Reflective Scepticism	'T'* (8) 'I often re-evaluate my experiences so that I can learn from them'	3,73	.83
	'S'* (9) 'I usually check the credibility of the source of information before making judgments'	3,86	.97
	'K'* (10) 'I usually think about the wider implications of a decision before taking action'	3,79	.87
	'U'* (11) 'I often think about my actions to see whether I could improve them'	4,00	.73

*Capital letters used by Sosu (2013) while defining the Critical Thinking Disposition Scale is identified by numbers in the translated Norwegian version.

Cronbach's alpha was used to measure the reliability of the construct. The Cronbach's alpha coefficient of the subscales was .66 for '*critical openness*', .67 for '*reflective scepticism*', and .76 for the overall scale, respectively. The corrected item-total

correlations of CTDS-NO ranged from .27 to .65, which indicates good to very good discrimination. KMO was .81, exceeding the recommended value of .6 (Tabachnick and Fidell, 2014). 'Barlett's test of sphericity' reached statistical significance ($\leq$ .01) supporting the factorability of item-correlation. The two factors explained 41 % of the total variance.

CFA was conducted to test the overall goodness-of-fit of the two-dimension CTDS-NO with maximum likelihood estimates. All items loaded significantly ($\leq$ .001) to their construct dimension, and the results of CFA show a strong fit between the data (N = 182) and the measurement model. According to Brown (2015), fit indices can be broadly characterized into three categories: absolute fit, fit adjusting for model parsimony, and comparative or incremental fit. As each type of index provides different information about model fit, it is advised to consider and report at least one index from each category when evaluating the fit of a model. The overall 'goodness-of-fit' indices show strong support of contentions between CTDS (Sosu, 2013) and CTDS-NO. As presented in Table 2, the reasonability of the model is promising with strong appropriateness of a two-factor model (*p*=.093; SRMR .053). Parsimony correction indices indicate a good model fit (RMSEA=.040), showing similar index values as the original data set as to measure of approximate fit (CTDS; Sosu, 2013). Comparative fit indices evaluate the fit of a user-specified solution in relation to a more restricted, nested baseline model. Values (TLI=.94; CFI=.96; GFI=.95) indicate a good model fit, with values similar to the original data set. Both CTDS-NO and CTDS (Sosu, 2013) have similar Chi-square value (55.65 / 56.52) with values indicating a good fit factor dimension without any significant association (*p*=.093) between the subscales.

Table 2
'Goodness-of-fit' statistics for CFA

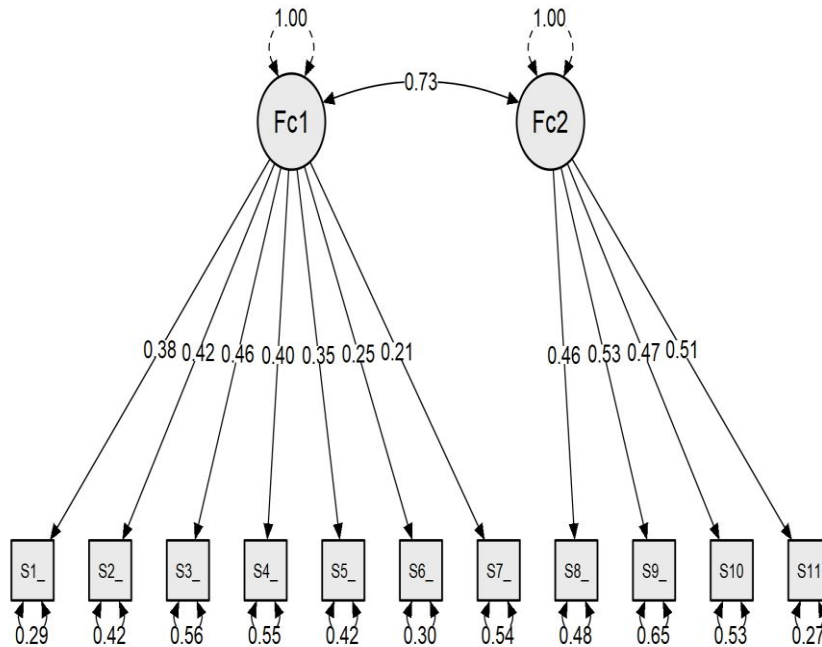|	X^2	df	GFI	TLI	CFI	RMSEA	SRMR
CTDS*				.91	.94	.051	.037
CTDS**				.96	.98	.038	.028
CTDS-NO	P=0,093	43	.95	.94	.96	.040	.053

* Sosu, 2013, initial specifications for the fit index
** Sosu, 2013, two-factor correlated errors model

Chi-square test

Model	X^2	df	p
Baseline model	343.220	55	
Factor model	55.652	43	0.093

Factor loadings and the path diagram for the Norwegian version of CTDS are presented in Figure 1. Items 2 to 4 and 8 to 11 load strong (> .4) to their dimension. Items 1 and 5 to 7 have a weaker loading (.21 - .38) to their dimension. Still, the two-dimension model shows a strong fit as measured by the fit indices, with a significant factor covariance (< .001) estimated at .73.

Model 1
Confirmatory factor analysis

DISCUSSION

This study presents a 'critical thinking disposition scale' (CTDS-NO) as a tool to measure the educational effectiveness of students' critical thinking disposition by exploring students' perceptions of their critical thinking. While the original questionnaire by Sosu (2013) was developed in English, the current study adapts and investigate the questionnaire's psychometric properties within a new linguistic group and cultural context. The 'goodness-of-fit' statistics for CFA suggest a factor structure that harmonises with the original factor structure of CTDS (Sosu, 2013), strengthening CTDS for cross-cultural use. The satisfying psychometric properties suggest high reliability. Internal consistency was assessed by Cronbach's alpha, with values indicating good internal consistency (Hair et al., 2014). The descriptive statistics reveal a tendency for ceiling effects concerning items loading in both dimensions. Responses with a positive skew towards the favourable end are, however, expected and considered normal. This situation often appears when a rating is used to evaluate students or staff (Streiner & Norman, 2008).

Although CTDS-NO show strong psychometric properties, which strengthens adaption of the original questionnaire for cross-cultural assessment, 'goodness-of-fit' indices

constitute only one aspect of model evaluation (Brown, 2015). Although model evaluation usually begins with the examination of these fit indices, it is equally important to examine a solution in terms of potential areas of localized strain and the interpretability and strength of the resulting parameter estimates. As regards to cross-cultural use, two other studies have previously explored the CTDS validity concerning respondents from countries other than the United Kingdom; Akin et al. (2015) and Yockey (2016). Akin et al. (2015) adapted and validated the CTDS (Sosu, 2013) into a Turkish version, which involved both translation and validation of the questionnaire. The results from CFA demonstrate that the 11 items loaded on the same factors as the original, showing strong psychometric properties implicating the instrument to be reliable for measuring critical thinking disposition and strengthening its cross-cultural use.   Unlike the Norwegian and Turkish adaptions, Yockey (2016) investigate the reliability, factor structure, and convergent validity of the CTDS (Sosu, 2013) on an ethnically diverse sample of participants from the United States by using the original developed form without cultural adaption adjustments. Yockey (2016) found that neither a one- or two-factor model optimally fit his data, with all fit indices falling outside the recommended guidelines for good model fit. However, similar to Sosu (2013), Yockey (2016) examines a two-factor correlated errors model, arguing in favour of this model due to common wording between items 1 and 2 of the CTDS. The two-factor correlated errors model shows improved model fit, with values falling within the criteria of acceptable fit. All though Yockey (2016) emphasizes the 'goodness-of-fit' indices, potential areas of localized strain are omitted from the assessment of the validity of the questionnaire. Adapting questionnaires from one country and culture to another necessitates both linguistic and cultural assessment, also where the written languages are highly comparable (Hambleton, 2005). Even though the original instrument is in English, the applicability of the instrument within other English-speaking countries is unknown. Adaption for cross-cultural assessment therefore needs to include both linguistic and cultural examination in addition to fit statistics.

While the current study found support for the two-dimension model, a few differences between the present work and previous adaptions and validations of the CTDS (Sosu, 2013) are worth of note. CFA reveals that the two dimensions in the Norwegian version of CTDS consist of the same items and with significant loadings to the respective dimensions. The model fit indices for the CTDS-NO scale indicates a good fit to the model presented by Sosu (2013). The model fit indices and the overall loadings in CTDS-NO show values exceeding Sosu's initial specifications for the fit index, and equal or near to values from Sosu's correlated errors model (2013). This confirms a good model fit for the CTDS-NO. Examination of a possible two-factor correlated errors model due to common wording between items 1 and 2 (Items N and C; Sosu, 2013) of the CTDS, as both Sosu (2013) and Yockey (2016) performed, was excluded due to no additional contribution to the model fit analysis of CTDS-NO.

Cronbach's alpha was used to measure the reliability of the construct, where values above .70 are considered as indicating an acceptable level of reliability (Cronbach, 1951; Streiner & Norman, 2008). However, as this is an adapted instrument, values above .60 are acceptable on subscales, with an overall scale above .70 (Village & Hood,

2017). Sosu (2013) has labelled the two CTDS dimensions '*critical openness*' and '*reflective scepticism*' with a high coefficient (α = .81) of internal consistency concerning the overall scale. The Cronbach's alpha coefficient of the CTDS-NO subscales were .66 for '*critical openness*', .67 for '*reflective scepticism*', and .76 for the overall scale. The reliability of the instrument is therefore within the values of an acceptable level. Items 2-4 and 8-11 (Items K, S, T, U; Sosu, 2013) loads strong (>.4) to their dimension. Items 1 and 5-7 (Items N, O, J, E; Sosu, 2013) have a weaker loading (.21-.38) to their dimension. Still, the two-dimension model shows a strong fit as measured by the fit indices (GFI = .95, TLF = .94, CFI = .96, RMSEA = .040, SRMR = .053). Table 2 compares the fit statistics of CTDS (Sosu, 2013) and CTDS-NO. Overall findings demonstrate that CTDS-NO has satisfying psychometric properties and high reliability.

When developing CTDS, Sosu (2013) accepts item loadings above .30 as acceptable loadings. Although the overall item loadings are stronger for the CTDS than the adapted CTDS-NO, all items except items 6 and 7 (Items J and E; Sosu, 2013) load above .30 in CTDS-NO and are defined as acceptable loadings, according to Sosu (2013). Differences in item loading values may be related to different student populations used; the students represented different educational programmes and different levels within the programmes. The questionnaire is demanding, as it requires metacognitive processing. With respondents being first year undergraduate students, there is an expected discrepancy regarding students' reflection level compared to Sosu's (2013) respondents, which contained both undergraduate and graduate students. There is, however, a question of suitability concerning items 6 and 7 within the dimensions presented. Items 6 and 7 load to the dimension labelled '*critical openness'.* Both items load weak to their dimension, with item 6 at .25 and item 7 at .21, indicating that their suitability may fit stronger elsewhere. Analysing the linguistic content of both item 6 *'it is important to understand other people's viewpoint on an issue'* and item 7 *'it is important to justify the choices I make'*, in both the English and Norwegian form, raises the possibility of a third dimension in the form of possible single items. Item 6 concerns the exploration of another person perspective, *'it is important to understand other people's viewpoint…'*, while item 7 focuses on the defence of action, '*It is important to justify the choices …'.* Compared to the content of items 1 to 5, where the students are asked to reflect on their own opinions or position, e.g. *'I usually try to think …', '… why I do things', '… information about myself'*, it is reasonable to question whether items 6 and 7 belong within the "*critical openness*" dimension. However, both exploring another person's point of view and justifying choices or action may be within the "*critical openness*" dimension. Both perspectives demand a critical reflection on the underlying understanding of one's disposition. The dimension has a broad application, which allows the items to have different variables within the same dimension.

Cronbach's alpha coefficient is .66 for the '*critical openness*' dimension when items 6 and 7 are included in the dimension. Item statistics show that removing item 6 from the dimension will reduce the reliability of the scale. This confirms that the item has its affiliation within the dimension even though the linguistic content also allows for other interpretations (Hambleton & De Jong, 2003; Hambleton, 2005). Removing item 7 from

the dimension will, however, increase the total scale reliability slightly with a Cronbach alpha coefficient of .67, but the removal does not have an impact on the overall scale. Exploratory analysis identified item 7 with strong correlations (.789) outside the two-dimension solution, leading to a possible single item in the instrument. Still, neither the overall scale, nor the factor dimension showed change in value of significance when deleting item 7 from the dimension or the instrument. Model fit values became weaker when item 7 was deleted from the instrument, with decreased values on TLI and CFI, and increased values on RMSEA and SRMR. This indicates that CTDS-NO holds a higher validity by including item 7. The item therefore remained in the questionnaire, with a two-dimension structure as suggested by Sosu (2013).

Student's Awareness of Their Critical Thinking Disposition

CTDS-NO is an instrument that addresses the social work students' awareness of their critical thinking disposition. At the same time, the CTDS-NO visualises the students' perception through the two subscales '*critical openness*' and '*reflective scepticism*'. Descriptive statistics in this study show the students to be quite confident in their ability to think critically, even though they are just first year undergraduate students. The students reported high confidence in their ability to 'understand other people's viewpoint' (item 6) and self-reflect on 'their own actions' (item 11), with an overall belief above mean level on their 'critical openness' and 'reflective scepticism'. The result may indicate that this is a group with high competence within critical thinking. Based on international studies (Orszag, 2015), it is, however, more likely that the students' belief in their capability and their view of which skills and dispositions contribute to critical thinking are not proportional. Either way, increased awareness will affect their skills. Using the questionnaire within social work programmes will not only help address the students' awareness and development towards critical thinking in professional practice, it may also be useful for measuring whether the teaching contributes to the development of critical thinking.

LIMITATIONS

All though the overall findings demonstrate that CTDS-NO has satisfying psychometric properties and high reliability, the current study has some limitations. First, sample size may be held as a limitation as there is little agreement on how large a sample should be. In small samples the correlation among items are less reliable, and where factors obtained from small data sets do not generalise as well as those derived from larger samples. Some suggest that is not the overall sample size that is of concern, rather than the ratio of participants to items. This study is based on a sample size ratio of 16:1, exceeding recommended cases for each item to be factor analysed (Tabachnick & Fidell, 2014). Nonetheless, increasing sample size could strengthen both item and factor correlations.

Second, the CTDS is a questionnaire that requires metacognitive principals and awareness. Sosu (2013), Akin et al. (2015), Yochey (2016) and this present study all use undergraduate students as participants. Knowing that metacognition develop and expand through practice (Hartman, 2001), there is reason to believe that participants with

different backgrounds and other educational level would obtain different values using the same form. Hence, whether another sample group would strengthen the use of the form needs further exploration.

Third and last, it is possible that differences in the structure and emphases of educational systems could result in differences regarding the conceptualisation of the critical thinking disposition construct. A cross-cultural assessment before use within other countries is therefore recommended.

CONCLUSION

Overall findings demonstrate that the adapted and translated Norwegian version of CTDS has satisfying psychometric properties and high reliability. Moreover, it may be a valuable tool for measuring educational effectiveness on students' critical thinking disposition during their education. With participants consisting of first year undergraduate social work students from bachelor programmes in Norway, CTDS-NO should also be a valuable tool for students within higher educational levels. Nevertheless, the CTDS-NO needs further validation and testing in other contexts with larger samples. Further research should also explore the item correlations and dimension consistency.

REFERENCES

Akin, A., Hamedoglu, M.A, Arslan, S., Akin, Ü., Celik, E., Kaya, C. & Arslan, N. (2015). The adaption and validation of the Turkish version of the critical thinking disposition scale (CTDS). *The International Journal of Educational Researchers, 6*(1), 31-35.

Boot, R.L., & Reynolds, M. (1983). *Learning and experience in formal education. Manchester monographs*. Manchester: Manchester University Press.

Brookfield, S.D. (2012). Critical theory and transformative learning. In: E.W. Taylor & P. Cranton (Eds.). *The handbook of transformative learning: Theory, research and practice* (p.131-146). Jossey-Bass.

Brookfield, S.D. (2014). Repressive tolerance and practice of adult education. International *Journal of Adult Vocational Education and Technology, 5*(2), 15-28.

Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. Second Edition. Guilford.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Ennis, R.H. (2008). Nationwide testing of critical thinking for higher education: Vigilance required. *Teaching Philosophy*, *31*, 1-26.

Field, A. (2017). *Discovering statistics using IBM SPSS statistics*. 5th edition. Sage.

Guillemin F, Bombardier C, Beaton D (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol, 46*, 1417-1432.

Hair, J.F., Black W.C., Babin, B.J., & Anderson, R.E. (2014). *Multivariate data analysis*, 7th edition. Pearson New International. Pearson.

Hambleton, R.K. (2005). Issues, designs, and technical guidelines for adapting test into multiple languages and cultures. In: Hambleton, R.K, Merenda, P.F. & Spielberger, C.D. (Ed.). *Adapting educational and psychological tests for cross-cultural assessment.* Psychology Press.

Hambleton, R.K. & De Joung, J.H.A.L. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing, 20,* 127-134.

Harkness, J.A., Mohler, P.P, & Van de Vijver, F.J.R. (2003). Comparative research. In Harkness, J.A., Van de Vijver, F.J.R., & Mohler, P.P. (Eds.), *Cross-cultural survey methods.* John Wiley & Sons.

Hartman, H.J. (2001). *Developing students' metacognitive knowledge and skills*. In Hartman, H.J. (Red). Metacognition in learning and instruction: Theory, research and practice. Kluwer Academic.

Hooks, B. (2010). *Teaching critical thinking: Practical.* Wisdom. Routledge.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods 6*(1), 53–60.

Hu, L.T. and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Meadows, K. (2012). *User guide. the diabetes health profile* (DHP-1 and DHP-18).

Melguizo, T., Zamorro, G., Velasco, T., & Sanchez, F.J. (2017). The Methodological Challenges of Measuring Student Learning, Degree Attainment, and Early Labor Market Outcomes in Higher Education. Journal of Research on Educational Effectiveness, *10*(2), 424-448, DOI: 10.1080/19345747.2016.1238985

Merriam, S.B., & Bierema, L.L. (2014). *Adult learning: Bridging theory and practice*. Jossey-Bass.

Nugroho, P.B., Nusantara, T., As'ari, A.R. Sisworo, Hidayanto, E., & Susiswo (2018). Critical Thinking Disposition: Students Skeptic in Dealing with Ill-Logical Mathematics Problem. *International Journal of Instruction*, *11*(3), 635-648.

Polit, D.F. & Beck, C.T. (2016). Designing and conducting quantitative studies to generate evidence for nursing. In: Polit D.F. & Beck C.T. (Eds.). *Nursing research*. 10th ed. Wolters Kluwer Health, p. XIV, 802 s.

Poondej, C. & Lerdpornkulrat, T. (2015). The reliability and construct validity of the critical thinking disposition scale. *Journal of Psychological and Educational Research, 23*(1), 23-26.

Rahdar, A., Pourghaz, A., & Marziyeh, A. (2018). The impact of teaching philosophy of children on critical openness and reflective skepticism in developing critical thinking and self-efficacy. *International Journal of Instruction*, *11*(3), 539-556.

Sireci, S.G. (2005). Using bilinguals to evaluate the comparability of different language versions of a test. In: Hambleton, R.K, Merenda, P.F. & Spielberger, C.D. (Ed.). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Psychology Press.

Sireci, S.G., Patsula, L., & Hambleton, R.K. (2005). Statistical methods of identifying flaws in the test adaption process. In: Hambleton, R.K, Merenda, P.F. & Spielberger, C.D. (Ed.). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Psychology Press.

Sosu, E.M. (2013). The development and psychometric validation of a critical thinking disposition scale. *Thinking Skills and Creativity, 9*, 2013, 107-119.

Streiner, D.L. & Norman, G.R. (2008). *Health measurement scales: A practical guide to their development and use.* 4th Edition. Oxford University Press.

Suryanti, N. & Nurhuda (2021). The effect of problem-based learning with an analytical rubric on the development of students' critical thinking skills. *International Journal of Instruction*, *14*(2), 665-684.

Tabachnick, B.G. & Fidell, L.S. (2014). *Using multivariate statistics. 6th edition. pearson new international edition*. Pearson/Allyn and Bacon.

Van de Vijver, F.J.R. & Hambleton, R.K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89-99.

Van De Vijver, F.J.R & Poortinga, Y.H. (2005). Conceptual and methodological issues in adapting tests. In: Hambleton, R.K, Merenda, P.F. & Spielberger, C.D. (Ed.). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Psychology Press.

Village, A. & Hood, R.W. (2017). *Research in the social scientific study of religion, volume 28*. Doi: https://doi.org/10.1163/9789004348936

Watkins, M.W. (2006). Determining parallel analysis criteria. *Journal of Modern Applied Statistical Methods, 5*(2), 344-346.

Ware, J.E. (1995). The status of health assessment 1994. *Annu Rev Public Health, 16*, 327-354.

Yockey, R.D. (2016). Validation study of the critical thinking dispositions scale: A brief Report. *North American Journal of Psychology, 18*(1), 101-106.