IJPES
www.ijpes.com
International Journal of Psychology and Educational Studies
ISSN: 2148-9378

Investigating Psychometric Properties of the Turkish Version of Sosu Critical Thinking Disposition Scale: Evidence from Two Independent Samples

Ali ORHAN[1]

[1] *School of Foreign Languages, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye* *0000-0003-1234-3919*

ARTICLE INFO

Article History
Received 16.08.2022
Received in revised form 21.11.2022
Accepted 24.02.2023
Article Type: Research Article

ABSTRACT

This study aimed to examine the psychometric properties of the Turkish version of Sosu Critical Thinking Disposition Scale (TV-CTDS). The data were gathered from two independent samples consisting of undergraduate and graduate students from four different universities in the Black Sea region of Turkey. After the translation procedure, the psychometric properties of the TV-CTDS, including factor structure, measurement invariance, convergent validity, discriminant validity, and reliability, were investigated. The results of the study revealed that the two-factor model (critical openness and reflective skepticism) of the TV-CTDS presented a better fit to the data than the one-factor model, and two-factor model is invariant across different samples (undergraduate and graduate students) and genders. Also, the TV-CTDS showed convergent validity and discriminant validity across two different samples, and its internal consistency was adequate. Therefore, it can be said that the TV-CTDS can be used by educators, researchers, and professionals to assess undergraduate and graduate students' critical thinking dispositions in Turkey.

Keywords:
Critical thinking dispositions, Sosu Critical Thinking Disposition Scale, psychometric properties, measurement invariance, scale adaptation

1. Introduction

Critical thinking (CT) can be regarded as one of the most essential attributes for individuals to be successful in the 21st century (Huitt, 1998). Therefore, promoting CT at all levels of education is seen as one of the central goals of education systems in many countries today (Stassen et al., 2011; Stedman & Adams, 2012) because CT is essential not only for the academic achievement of individuals at school (Tican & Taşpınar, 2015) but also for their lifelong learning (Halpern, 2003).

CT is a functional, reasonable, and reflective thinking process which focuses on making a judgment about the accuracy and quality of knowledge (Beyer, 1995) and is employed by individuals to decide what to do or what to believe (Ennis, 1991). Ruggerio (1990) simply defined CT as a logical and rational thinking process about knowledge, information, and ideas. Individuals think about ideas, information, and knowledge logically and investigate the source of them, evaluate their reliability, and test their validity thanks to CT (Paul, 1990). In other words, individuals can obtain rational, useful, and true information through CT (Lewis & Smith, 1993) which can be seen as an important guard that can defend individuals against false information in this century (Epstein & Kernberger, 2012). Therefore, it can be said that CT is an important attribute that each and every individual should possess.

CT can be categorized into two dimensions as skills and dispositions. In other words, CT is not only about cognitive skills such as deduction, making inferences, and interpretation, but also about some dispositions or

[1] Corresponding author's address: School of Foreign Languages, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye
e-mail: ali_orh_an@hotmail.com
Citation: Orhan, A. (2023). Investigating psychometric properties of the Turkish version of sosu critical thinking disposition scale: Evidence from two independent samples. *International Journal of Psychology and Educational Studies*, *10*(2), 348-359. https://dx.doi.org/10.52380/ijpes.2023.10.2.1017

attitudes towards using these cognitive skills (Ennis, 1987a; Paul, 1990). Critical thinking dispositions (CTD) that can be defined as a tendency or willingness to use critical thinking skills (CTS) (Pascarella & Terenzini, 1991) are vital for being a successful critical thinker (Profetto-McGrath, 2003) because individuals with high CTS but low CTD will probably have problems using CTS in the right situations (Siegel, 1988). Also, having only high CTD does not make the individual a good critical thinker. Therefore, we can say that both CTS and CTD are vital for individuals because without dispositions towards using CTS, individuals may never use these skills; conversely, without having high CTS, strong dispositions will be useless for individuals (Profetto-McGrath et al., 2003).

CT should be regarded as a collective skill and it composes of some sub-skills (Fisher, 2001). When the literature is examined, many classifications regarding CTS can be seen (e.g., Ennis et al., 2005; Facione, 2000, Watson & Glaser, 1994). Like CTS, several classifications regarding CTD have also been described in the literature. It can be said that a Delphi study by APA (1990) provided the most comprehensive classification of CTD with 19 broad sub-dimensions. Facione and Facione (1992) reduced these 19 sub-dimensions to seven and developed the California Critical Thinking Disposition Inventory (CCTDI) to measure individuals' CTD. Although the CCTDI, which has 75 items and seven subscales, can be considered as the first instrument to measure CTD and has been widely used by many researchers since its development, it has some problems. Some researchers (Bondy et al., 2001; Kakai, 2003; Walsh & Hardy, 1997; Walsh et al., 2007) reexamined the psychometric qualities of the CCTDI and revealed that there are inconsistencies in the factor structure of the scale. Also, in his reliability generalization meta-analysis study aiming to estimate a general alpha value for the CCTDI and its subscales, Orhan (2022) concluded that while the general alpha estimates of the subscales ranged from 0.56 to 0.74, it was found to be 0.83 for the total CCTDI. Therefore, it can be said that the reliability coefficients of most of the subscales were below the desired level of 0.70 (Nunnally & Bernstein, 1994). It is recommended to develop a shorter version of the CCTDI because of these inconsistencies in the factor structure of the CCTDI that emerged in cross-validation studies (Walsh et al., 2007).

The UF/EMI Critical Thinking Disposition Instrument (CTDI) developed by Irani et al. (2007) is another scale to measure high school students' CTD. The conceptualization of the CTDI derives from the work of Facione (1990) and the CCTDI. Irani et al. (2007) examined the data obtained from the factor analysis of the CCTDI and aimed to develop a new instrument, namely, the CTDI. The CTDI has 25 items and 3 subscales which are engagement, maturity, and innovativeness. The reliability values of the subscales were 0.90, 0.78, and 0.79 for engagement, maturity, and innovativeness, respectively. The Sosu Critical Thinking Disposition Scale (CTDS) developed by Sosu (2013) is another important instrument to measure undergraduate and graduate students' CTD.

The CTDS was developed by Sosu (2013) after reviewing several taxonomies of CTD. The CTDS was reduced to 11 items under two factors from an initial 24 item-pool after exploratory factor analysis (EFA) with sample 1 (467 undergraduate and graduate students). All items had loadings of 0.35 or above. Then, Sosu (2013) conducted a confirmatory factor analysis (CFA) with sample 2 (371 undergraduate and graduate students) to evaluate the two-factor model obtained after EFA. After CFA, it was seen that two-factor model showed a good model fit for undergraduate sample ($MLMX^2_{(42)}$=56.72, p=0.00; X^2/df-ratio=1.35; RMSEA=0.05; CFI=0.93; TLI=0.91; SRMR=0.05) and for graduate sample ($MLMX^2_{(42)}$=76.96, p=0.00; X^2/df-ratio=1.83; RMSEA=0.05; CFI=0.94; TLI=0.92; SRMR=0.05) with the residual covariance between two items from the critical openness (CO) subscale. He also tested measurement invariance (MI) across undergraduate and graduate samples with the data obtained from sample 2 and found that the two-factor model was same for both of the groups and the measurement model was invariant across these two groups. The CTDS, which uses a 5-point Likert scale, has two subscales which are CO and reflective skepticism (RS). The CO subscale has seven items (e.g., "I am often on the lookout for new ideas") and the RS subscale has four items (e.g., "I often re-evaluate my experiences so that I can learn from them"). The CO subscale is about measuring the disposition of being open to new ideas and being critical while examining these new ideas. Individuals with a positive inclination toward the CO subscale are willing to modify their ideas, claims, and judgments in the light of strong and convincing evidence. The RS subscale targets to measure willingness to learn from previous experiences and individuals with a high inclination toward the RS subscale treasure the constant questioning of evidence. The Cronbach's alpha value for the CTDS was 0.79 for sample 1 and 0.81 for sample 2.

Yockey (2016) reexamined the psychometric properties of the CTDS with 210 undergraduate students in the United States of America. Alpha coefficients were calculated as 0.67 and 0.71 for CO and RS subscales, respectively. Yockey (2016) also investigated the convergent validity (CV) of the CTDS by calculating the correlation value between the CTDS and Need for Cognition Scale (NFCS) and found a significant relationship between the CTDS and NFCS scores. He examined one-factor and two-factor structure of the CTDS both with and without correlated errors between two items using CFA and found out that both of the models with correlated errors between two items presented a good model fit for the data. Although no significant change in X^2 was observed between one and two-factor structures, he suggested using the CTDS with more parsimonious one-factor model because of the very high correlation between two subscales in two-factor without correlated errors and two-factor with correlated errors models (r=0.82 and 0.99).

The CTDS was reexamined by Thomas and Hayes (2021) with a sample group of 558 undergraduate and graduate students from the United States of America. Both two-factor model identified by Sosu (2013) and some other possible alternatives including one-factor and three-factor models were tested. Also, Thomas and Hayes (2021) investigated the CV of the CTDS by calculating the correlation estimates between the CTDS, NFCS and Flexible Thinking Scale (FTS). After exploratory structural equation modeling procedure, it was found out that two-factor structure showed a good model fit (RMSEA=0.05; CFI=0.96; TLI=0.94; SRMR=0.02). However, after an investigation of the factor loadings, it was seen that two items (item 1 and 3) loaded simultaneously onto both factors. Therefore, they excluded these two items and examined two-factor model again and found out that a reduced two-factor model showed a better fit to the data (RMSEA=0.05; CFI=0.97; TLI=0.95; SRMR=0.02). The correlation values revealed a significant relationship between the CTDS, NFCS, and FTS scores providing evidence for the CV of the CTDS. Reliability coefficients were calculated as 0.72 and 0.80 for CO and RS subscales, respectively.

Bravo et al. (2020) examined the psychometric qualities of the Spanish version of the CTDS (SV-CTDS) with 1064 students who are studying in the last two years of high school (N=703) and at university (N=361) in Spain. After translation procedure of the scale into Spanish, two-factor model proposed by Sosu (2013), one-factor model, and bifactor model were tested using CFA. The CFA results showed that all three models showed an adequate model fit. However, Bravo et al. (2020) stated that one-factor model was the best model to represent the data and recommended it like Yockey (2016) because of the high correlation value between two subscales in two-factor model (r=0.93). Alpha estimates of the total scale were calculated as 0.76 [95% CI=0.72-0.80] for the sample of university students and 0.76 [95% CI=0.73-0.79] for the sample of high school students. Composite reliability (CR) values were 0.84 and 0.81 for university students and high school students, respectively. Also, MI of the one-factor model was assessed across gender groups and the results revealed that one-factor model is invariant and it can be used across different genders.

Luiz et al. (2021) carried out a study with a sample group of 179 undergraduate students in Brazil to adapt the CTDS into Brazilian Portuguese and to investigate psychometric qualities of the Brazilian Portuguese version of the CTDS (BPV-CTDS). After translation procedure of the scale into Brazilian Portuguese, EFA was conducted to reveal the factor structure of the BPV-CTDS. EFA indicated the presence of two factors with total variance of 44.88%. However, parallel analysis of the items and factors showed that there was inadequacy in this structure and Luiz et al. (2021) conducted another EFA fixing the solution in a single factor and found out that 11 original items had good factor loadings onto the single factor (the lowest λ=0.49). Therefore, Luiz et al. (2021) stated that BPV-CTDS is better with one-factor model. As an evidence for the reliability of the BPV-CTDS, Omega coefficient was calculated and it was seen that the BPV-CTDS was reliable (Omega coefficient=0.79; 95%CI=0.74-0.83).

Gerdts-Andresen et al. (2022) investigated psychometric qualities of the Norwegian version of the CTDS with 182 undergraduate students in Norway. After translation procedure of the scale into Norwegian, they tested the traditional two-factor model by CFA and found that two-factor structure fit the data well (RMSEA=0.04; CFI=0.96; TLI=0.94; SRMR=0.05). Cronbach's alpha values were 0.66, 0.67, and 0.76 for CO, RS, and total scale, respectively. Akın et al. (2015) conducted a study with 212 undergraduate students in Turkey to adapt the CTDS into Turkish and to examine the validity and reliability of it. After translating the scale into Turkish, factor structure of the Turkish version of the CTDS (TV-CTDS) was investigated by CFA and Cronbach's alpha value was used to examine its reliability. The results of CFA revealed that two-factor structure presented a good model fit (X^2=53.24, df=40, RMSEA=0.04, GFI=0.96, CFI=0.97, SRMR=0.04). Factor loadings of the items

were between 0.27 and 0.59 for CO subscale and they ranged from 0.55 to 0.76 for RS subscale. Reliability estimates were 0.68, 0.75, and 0.78 for CO, RS, and total scale, respectively.

Previous literature indicates that the CTDS has been widely used by many researchers to determine individuals' CTD since the day it was developed. The CTDS has been translated into many languages such as Spanish (Bravo et al., 2020), Brazilian Portuguese (Luiz et al., 2021), Norwegian (Gerdts-Andresen et al., 2022), and Turkish (Akın et al., 2015). Besides, there are other studies investigating the psychometric properties of the original CTDS (Thomas & Hayes, 2021; Yockey, 2016). After factor analytic procedures, some studies revealed that the original two-factor model proposed by Sosu (2013) showed the best model fit (Akın et al., 2015; Gerdts-Andresen et al., 2022; Thomas & Hayes, 2021) although some studies concluded that the one-factor model had better psychometric properties for the CTDS (Bravo et al., 2020; Luiz et al., 2021; Yockey, 2016).

In their Turkish adaptation study, Akın et al. (2015) just tested the two-factor model of the TV-CTDS by CFA and calculated the reliability estimates. They did not provide any evidence for MI, CV, and discriminant validity (DV). So, this study aimed to investigate the psychometric qualities of the TV-CTDS including factor structure, MI, CV, DV, and reliability. In this study, firstly two competitive factor structures (the original two-factor structure and the one-factor structure) were investigated by CFA. Then, MI across two independent groups and two genders was examined by multi-group CFA. After that, Cronbach's alpha estimates were calculated to investigate the reliability. Also, the CV and DV of the scale were investigated. It can be said that this study is different from the first Turkish adaptation study by Akın et al. (2015) because this current study is more comprehensive and provides evidence for the original two-factor model emerged in the test manual by Sosu (2013) and the one-factor model identified in some previous studies (Bravo et al., 2020; Luiz et al., 2021; Yockey, 2016), MI of the scale across two independent groups and two genders, CV and DV, and reliability.

2. Methodology

2.1. Translation Process of the CTDS

The original CTDS with 11 items was translated from English into Turkish by two English instructors working in the School of Foreign Languages (SFL) at a state university. After that, two versions of the translation were compared with an expert who has high English language proficiency by selecting the best alternative for each item and the best version was agreed on. Then, TV-CTDS was back-translated into English by another English instructor working in the SFL at a state university and had no idea about the original scale. After comparing the back translated version with the original scale, it was seen that the scale was accurately translated into Turkish. Finally, the TV-CTDS was checked by a Turkish teacher to ensure the accuracy of Turkish wording and grammar.

2.2. Participants

The data for this current study were collected from two independent samples (sample 1=516 and sample 2=249) consisting of undergraduate and graduate students who are studying at four different universities in the Black Sea region of Turkey in 2021-2022 academic year. I used the data gathered from two independent samples to reduce the possibility of sample specific bias. Also, using two independent samples made it possible to investigate the MI of the TV-CTDS across undergraduate and graduate students.

Sample 1 included 516 undergraduate students studying in different disciplines. 294 of these students were female (57%) and 222 of them were male (43%). The mean age of the undergraduate students was 19.95 (SD=1.26) ranging from 18 to 23.

Sample 2 consisted of 249 graduate students who are pursuing their master's or Phd degrees in different disciplines. 129 of these students were female (52%) and 120 of them were male (48%). The mean age of the graduate students was 23.80 (SD=1.00) ranging from 22 to 27.

2.3. Data Collection

The data were gathered in the regular scheduled face-to-face lessons with the help of the instructors of the lessons in the 2021-2022 academic year. Privacy and confidentiality issues were shared with the students and

they were informed about the study and their right to withdraw anytime they want. The scale was completed in approximately 15-20 minutes.

2.4. Data Analysis

The psychometric qualities of the TV-CTDS were investigated in terms of its factor structure, MI across two independent samples and gender, internal consistency, CV, and DV. First, the original two-factor structure of the TV-CTDS with 11 items proposed by Sosu (2013) and the one-factor structure were examined by CFA with the maximum likelihood estimation method. The fit of the models to the data was examined using some goodness-of-fit indices like the chi-square statistic, the Root Mean Squared Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the standardized root mean squared residuals (SRMR). TLI and CFI values over 0.90 show acceptable fit while values over 0.95 represent good fit (Hu & Bentler, 1999; Marsh et al., 2004). Also, RMSEA values should be lower than 0.05 to say good fit, while values between 0.05 and 0.08 show acceptable fit (Hu & Bentler, 1999; Kline, 2005). With respect to SRMR, values below 0.08 and 0.10 indicate good and acceptable fit, respectively (Hu & Bentler, 1999; Kline, 2005).

Second, a multi-group CFA was performed to have evidence of MI across two independent samples (undergraduate and graduate students) and gender. While performing multi-group CFA, firstly the configural model was tested. In the configural model, parameters are not constrained and they are all freely estimated for each group. After that, the metric invariance was tested. In this model, factor loadings were constrained across groups. Lastly, the intercepts of the groups are constrained and the scalar invariance was tested.

Third, the internal reliability of the TV-CTDS was examined by calculating Cronbach's alpha estimates. Fourth, the CV of the TV-CTDS was investigated using the average variance extracted (AVE) and CR values for the subscales. AVE values over 0.50 (Hair et al., 2010) and CR values over 0.60 (Fornell & Larcker, 1981) indicate a satisfactory CV. Fifth, the DV of the TV-CTDS was examined by comparing each subscale's AVE values with their respective shared variance estimates. The shared variance can be calculated by the square of the correlation coefficient and when each subscale's AVE is greater than its squared correlations with other factors, we can say that DV exists (Fornell & Larcker, 1981). Mplus version 6.0 (Muthen & Muthen, 2010) and SPSS version 23 were used to analyze the data.

2.5. Ethical

Ethical committee approval was obtained from Zonguldak Bülent Ecevit University (Date: 08.02.2022, No: 130948).

3. Findings

3.1. Confirmatory Factor Analysis

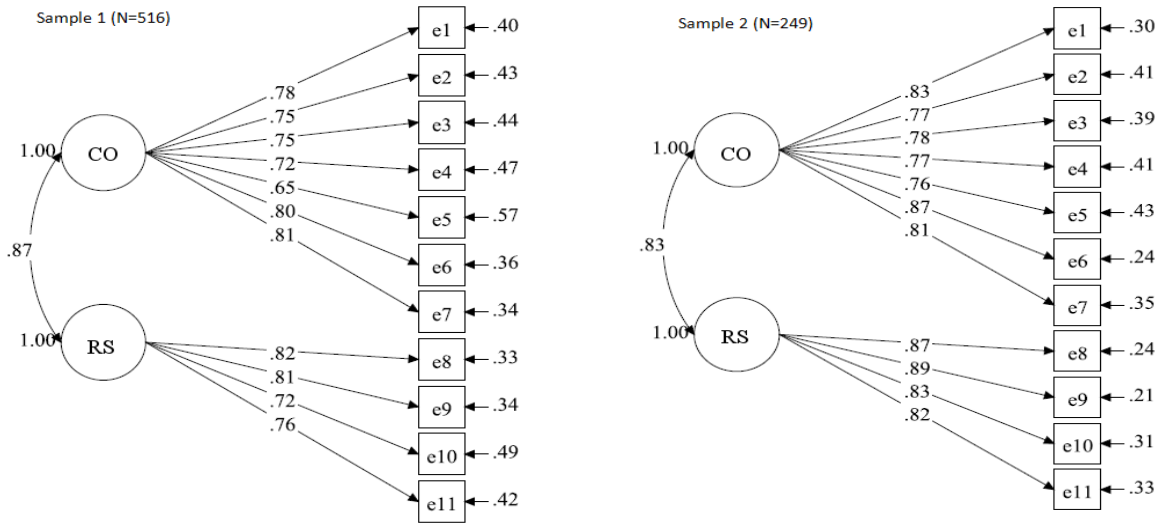
The original two-factor model of the TV-CTDS with 11 items proposed by Sosu (2013) and the one-factor model were evaluated by CFA in two independent samples. As it is shown in Table 1, CFA results showed that the two-factor model presented a good model fit for sample 1 ($X^2_{(43)}$=151.927, $p<0.05$; X^2/df-ratio=3.53; RMSEA=0.07; CFI=0.96; TLI=0.95; SRMR=0.02) and for sample 2 ($X^2_{(43)}$=85.198, $p<0.05$; X^2/df-ratio=1.98; RMSEA=0.06; CFI=0.98; TLI=0.97; SRMR=0.02). However, the one-factor model did not show an adequate model fit for both samples. It can be said that the two-factor structure originally proposed by Sosu (2013) showed a better model fit than the one-factor structure for the two independent samples of this study. Therefore, I went on the further analysis with two-factor structure of the TV-CTDS.

Table 1. *Goodness-of-fit Indices Regarding Two-factor and One-factor Model*

		X^2	df	RMSEA	CFI	TLI	SRMR
Sample 1 (N=516)	Model 1: two-factor model	151.927*	43	0.070	0.967	0.958	0.029
	Model 2: one-factor model	272.749*	44	0.100	0.932	0.914	0.041
Sample 2 (N=249)	Model 1: two-factor model	85.198*	43	0.063	0.980	0.974	0.028
	Model 2: one-factor model	243.477*	44	0.135	0.904	0.880	0.049

Note: *p<0.05

Path diagrams for both of the samples can be seen in Figure 1. As shown in Figure 1, factor loading of the items ranged from 0.65 to 0.82 and from 0.76 to 0.89 for sample 1 and 2, respectively.



Note: All of the standardized factor loadings and covariances are statistically significant (p<0.01); RS=Reflective skepticism; CO=Critical openness

Figure 1. *CFA Results for the TV-CTDS*

3.2. MI Using Multi-group CFA Across Two Samples

A multi-group CFA was performed to have evidence of MI to assure that the construct identities are same across two samples (undergraduate and graduate students) and the results were shown in Table 2.

Table 2. *MI Results Across Two Independent Samples for the Two-factor Model*

	X^2	df	ΔX^2	Δdf	p	RMSEA	CFI	TLI	SRMR
Sample 1 (N=516)	151.927*	43	-	-	-	0.070	0.967	0.958	0.029
Sample 2 (N=249)	85.198*	43	-	-	-	0.063	0.980	0.974	0.028
Configural invariance	237.125*	86	-	-	-	0.068	0.972	0.964	0.029
Metric invariance	243.364*	95	6.239	9	0.71	0.064	0.973	0.968	0.035
Scalar invariance	257.142*	104	13.778	9	0.13	0.062	0.972	0.970	0.037

Note: ΔX^2=difference in X^2; Δdf= difference in df; *p<0.05

The fit statistics for the baseline model were good ($X^2_{(86)}$=237.125, p<0.05; X^2/df-ratio=2.75; RMSEA=0.06; CFI=0.97; TLI=0.96; SRMR=0.02) which indicated configural invariance. Also, the fit statistics showed that the second model presented a satisfactory model fit ($X^2_{(95)}$=243.364, p<0.05; X^2/df-ratio=2.56; RMSEA=0.06; CFI=0.97; TLI=0.96; SRMR=0.03). The unconstrained model was not significantly different from the second model in which factor loadings were constrained which indicated metric invariance (p>0.05). Besides, the fit statistics of the third model were good ($X^2_{(104)}$=257.142, p<0.05; X^2/df-ratio=2.47; RMSEA=0.06; CFI=0.97; TLI=0.97; SRMR=0.03) and the second model was not significantly different from the third one suggesting scalar invariance (p>0.05). Therefore, it can be said that two-dimensional model is invariant and can be generalized across two groups (undergraduate and graduate students).

3.3. MI Using Multi-Group CFA Across Two Genders

A multi-group CFA was performed with the combined data (N=765) obtained from sample 1 and sample 2 to provide evidence of MI and assure that the construct identities are same across two genders and the results were shown in Table 3.

Table 3. *MI Results Across Two Genders for the Two-factor Model*

	X^2	df	ΔX^2	Δdf	p	RMSEA	CFI	TLI	SRMR
Male (N=342)	123.798*	43	-	-	-	0.074	0.964	0.955	0.032
Female (N=423)	124.841*	43	-	-	-	0.067	0.967	0.957	0.035
Configural invariance	248.639*	86	-	-	-	0.070	0.966	0.956	0.033
Metric invariance	262.169*	95	13.530	9	0.140	0.068	0.965	0.959	0.045
Scalar invariance	269.225*	104	7.056	9	0.631	0.064	0.965	0.963	0.050

Note: ΔX^2=difference in X^2; Δdf= difference in df; *p<0.05

The original two-factor structure of the CTDS with 11 items fit the data well for male participants ($X^2_{(43)}$=123.798, p=0.00; X^2/df-ratio=2.87; RMSEA=0.07; CFI=0.96; TLI=0.95; SRMR=0.03). Similarly, two-factor

structure showed satisfactory model fit for female participants ($X^2_{(43)}$=124.841, p=0.00; X^2/df-ratio=2.90; RMSEA=0.06; CFI=0.96; TLI=0.95; SRMR=0.03). The fit statistics for the baseline model were satisfactory ($X^2_{(43)}$=248.639, p<0.05; X^2/df-ratio=2.89; RMSEA=0.07; CFI=0.96; TLI=0.95; SRMR=0.03) which indicated configural invariance. Also, the fit statistics indicated that the second model showed a satisfactory model fit ($X^2_{(95)}$=262.169, p<0.05; X^2/df-ratio=2.75; RMSEA=0.06; CFI=0.96; TLI=0.95; SRMR=0.04). The unconstrained model was not significantly different from the second model in which factor loadings were constrained which indicated metric invariance (p>0.05). Besides, the fit statistics of the third model were good ($X^2_{(104)}$=269.225, p<0.05; X^2/df-ratio=2.58; RMSEA=0.06; CFI=0.96; TLI=0.96; SRMR=0.05) and the second model was not significantly different from the third one suggesting scalar invariance (p>0.05). Therefore, it can be said that two-dimensional model is invariant and the same model can be used across different genders.

3.4. Convergent Validity

The CV of the TV-CTDS was examined by calculating AVE for each of the two TV-CTDS subscales. As shown in Table 4, AVE values were 0.56 and 0.60 for CO and RS subscales in sample 1, respectively, and they are 0.63 and 0.72 for CO and RS subscales in sample 2, respectively. Therefore, it can be said that AVE values were greater than the desired value, so CV exists. I also investigated the CV of the TV-CTDS by calculating each factor's CR. As reported in Table 4, CR values were 0.90 and 0.86 for the CO and RS subscales in sample 1, respectively while they are 0.92 and 0.91 for the CO and RS subscales in sample 2, respectively which indicated that TV-CTDS showed a satisfactory CV across the two independent samples (undergraduate and graduate students).

Table 4. *Results Regarding the CV of the TV-CTDS*

Factor	Item	Sample 1 (N=516)			Sample 2 (N=249)		
		SFL	AVE	CR	SFL	AVE	CR
Critical openness	1	0.777	0.569	0.902	0.835	0.639	0.925
	2	0.754			0.769		
	3	0.748			0.784		
	4	0.725			0.769		
	5	0.653			0.755		
	6	0.797			0.871		
	7	0.813			0.807		
Reflective skepticism	8	0.822	0.607	0.860	0.872	0.726	0.913
	9	0.813			0.889		
	10	0.717			0.828		
	11	0.759			0.816		

Note: AVE= average variance extracted; SFL=standardized factor loadings; CR= composite reliability

3.5. Discriminant Validity

The DV of the TV-CTDS was investigated by comparing each factor's AVE value with its shared variance estimates as recommended by Fornell and Larcker (1981). The correlation values between the CO and RS subscales were 0.763 and 0.770 and shared variances between these two factors were 0.582 and 0.592 for sample 1 and sample 2, respectively. All AVE values (CO=0.569; RS=0.607 for sample 1; CO=0.639; RS=0.726 for sample 2) were greater than the shared variance estimates between two factors (0.582 for sample 1; 0.592 for sample 2) for two independent samples except for the CO subscale for sample 1 although it is slightly lower. Therefore, it can be said that the DV of the TV-CTDS existed for both of the samples.

3.6. Internal Consistency

Internal consistency of two subscales of the TV-CTDS was estimated by Cronbach's alpha. As it is shown in Table 5, the reliability coefficients were 0.90 for CO, 0.85 for RS, and 0.92 for total scale for sample 1 and they were 0.92 for CO, 0.91 for RS, and 0.94 for total scale for sample 2. Besides, for the total sample, the reliability coefficients were found to be 0.91, 0.88, and 0.93 for CO, RS, and the total scale, respectively. Therefore, it can be said that internal consistency values of the CO and RS subscales can be seen as adequate (Nunnally & Bernstein, 1995).

Table 5. *Internal Consistency Values of the TV-CTDS*

	Sample 1 (N=516)	Sample 2 (N=249)	Total Sample (N=765)

	M	SD	α	M	SD	α	M	SD	α
Critical openness	3.47	0.87	0.90	3.69	0.95	0.92	3.54	0.90	0.91
Reflective skepticism	3.44	0.91	0.85	3.55	1.09	0.91	3.48	0.97	0.88
Total scale	3.45	0.83	0.92	3.62	0.96	0.94	3.51	0.88	0.93

4. Conclusion, Discussion and Recommendations

As the development of CT at all levels of education is seen as one of the central goals of education systems in many countries today (Stassen et al., 2011; Stedman & Adams, 2012), providing educators and researchers with validated instruments to determine CT is clearly important because determining individuals' CT in a reliable and valid way is vital to determine the effectiveness of any effort or educational activities to develop CT. So, this study aimed to investigate the psychometric qualities of the TV-CTDS to provide a validated and reliable instrument that can be used with Turkish undergraduate and graduate students to measure CTD.

In this study, firstly, two competitive factor structures (the original two-factor and one-factor structures) were investigated by CFA with the data obtained from two independent samples. The CFA results indicated that the two-factor model (CO and RS subscales) presented a better fit to the data than the one-factor model. This factor structure for the TV-CTDS identified in this current study is same as the factor model emerged in the test manual by Sosu (2013). In addition, Thomas and Hayes (2021) confirmed the two-factor model of the original CTDS proposed by Sosu (2013) in their study. The CTDS has been translated into many languages and is in widespread use worldwide. Most of these adaptation studies revealed that two-factor structure of the CTDS has better psychometric properties (Akın et al., 2015; Gerdts-Andresen et al., 2022). In their adaptation study with 212 undergraduate Turkish students, Akın et al. (2015) concluded that the two-factor model was well fit. However, Yockey (2016) reexamined the psychometric qualities of the original CTDS and tested two competitive factor structures (the original two-factor structure and the one-factor structure) and found that both of the structures with correlated errors between two items presented a satisfactory model fit for the data. Although both of the structures presented a good model fit, he suggested using the CTDS with more parsimonious one-factor structure. Bravo et al. (2020), who examined the psychometric qualities of the SV-CTDS, tested three competitive factor models and revealed that all three models showed adequate data fit. However, they recommended the one-factor model because of the high correlation estimate between two subscales in the two-factor model. Luiz et al. (2021) conducted a study to adapt the CTDS into Brazilian Portuguese and found out that the BPV-CTDS is better with one-factor model. It can be said that a huge part of previous research showed that the two-factor model of the CTDS has good psychometric properties (Akın et al., 2015; Gerdts-Andresen et al., 2022; Thomas & Hayes, 2021). Also, results of the other studies revealed that both the one-factor and two-factor models showed good psychometric properties although the one-factor model was recommended for use because of the high correlation estimate between two subscales in the two-factor model (Bravo et al., 2020; Luiz et al., 2021; Yockey, 2016). Therefore, it can be said that two-factor model for TV-CTDS that emerged in this study is confirmed by the previous studies.

MI of the TV-CTDS across two independent groups (undergraduate and graduate students) and two genders (female and male) were also examined by multi-group CFA in this study. Results of the multi-group CFA revealed that the two-factor model is invariant and can be generalized across the two different groups (undergraduate and graduate students) and the same model can be used across different genders. Sosu (2013) investigated the MI of the CTDS across two independent groups (undergraduate and graduate students) and concluded that the scale is invariant across these two groups. Bravo et al. (2020) also tested the MI of the one-factor SV-CTDS across genders and concluded that MI existed. Therefore, the results of this current study are confirmed by Sosu (2013) and Bravo (2020). It can be said that although the CTDS has been adapted into many different languages and its psychometric qualities have been reinvestigated by various researchers, only two previous studies have investigated the MI of the CTDS. Testing for MI which enables us to investigate whether participants from different groups understand and interpret the same measure similarly (Bialosiewicz et al., 2013) plays a crucial role in psychological research and it is a critical addition to statistical procedures while developing a measure or adapting it into different languages (Chan, 2011). When the literature on developing or adapting CT measures is investigated, it can be seen that testing for MI is mostly neglected. This can be shown as an important problem that can devalue the measure development or adaptation studies regarding CT although it is impossible to generalize this problem across all fields of scientific inquiry.

CV and DV of the TV-CTDS were investigated by using AVE, CR, and the correlation values in this current study. AVE and CR values indicated that TV-CTDS showed an adequate CV across two independent samples. Also, this current study showed that AVE values were greater than the shared variance estimate between two factors indicating the DV of the TV-CTDS existed for both samples. Bravo et al. (2020) calculated CR values for their one-factor model for two different samples and two genders but did not calculate the AVE values. CR values calculated in the study (0.84, 0.81, 0.92, and 0.91 for university students, high school students, female participants, and male participants, respectively) indicated that SV-CTDS showed an adequate CV like TV-CTDS. From the previous studies investigating the psychometric properties of the original CTDS and different language versions of it, it can be seen that the CV and DV of the CTDS have not been investigated in these studies. Only Bravo et al. (2020) calculated CR values for their one-factor model which can be seen as evidence for the CV of the scale. Therefore, it can be said that this study is important because it provides evidence for the CV and DV of the TV-CTDS.

In this study, internal reliability of the TV-CTDS was investigated by calculating alpha estimates for two independent samples and it was found that alpha values were between 0.85 and 0.94 indicating internal consistency of the TV-CTDS was adequate. Sosu (2013) calculated alpha coefficient of the CTDS as 0.79 for sample 1 and 0.81 for sample 2 in the test manual. The alpha values of SV-CTDS calculated by Bravo et al. (2020) ranged from 0.76 to 0.90. As an analysis of evidence of BPV-CTDS's reliability, Luiz et al. (2021) calculated Omega coefficient as 0.70 indicating internal consistency of the BPV-CTDS was obtained. The alpha coefficients of the TV-CTDS calculated by Akın et al. (2015) were 0.68, 0.75, and 0.78 for CO, RS, and total scale, respectively. The alpha values of the Norwegian version of the CTDS calculated by Gerdts-Andresen et al. (2022) were 0.66, 0.67, and 0.76 for CO, RS, and overall scale, respectively. Therefore, it can be said that this current study revealed similar alpha values with the previous adaptation and validation studies of the CTDS.

In short, following the translation procedure, the factor structure, MI, CV, DV, and internal reliability of the TV-CTDS were investigated in this study and it was concluded that the two-factor model (CO and RS) of the TV-CTDS showed a better fit to the data and this model is invariant across different samples (undergraduate and graduate students) and genders (female and male). Also, TV-CTDS showed an adequate CV and DV across two independent samples and its internal consistency was adequate. Therefore, it can be said that TV-CTDS can be used by educators, researchers, and professionals to assess CTD of undergraduate and graduate students in Turkey. Although the CTDS was translated into Turkish by Akın et al. (2015) before, this current study is more comprehensive than Akın et al.'s (2015) because they just tested the two-factor model of the CTDS by CFA and calculated the reliability estimates and they did not provide any evidence for MI, CV, and DV. Therefore, I can say that this study provides some important results on the psychometric qualities of the TV-CTDS and indicates that the TV-CTDS can produce valid and reliable scores with Turkish undergraduate and graduate students. Also, this study has made an important contribution to the CT literature by presenting some important additional evidence for the cross-cultural validity of the original CTDS developed by Sosu (2013) which has been translated and adapted into many different languages until now. Besides, this study has also contributed to the CT literature in Turkey since it was the first study to examine the psychometric qualities of the TV-CTDS comprehensively by providing evidence for the factor structure, MI, CV, DV, and internal reliability of the TV-CTDS.

5. Limitations and Implications for Future Studies

Even though this study is important to shed light on the psychometric properties of the TV-CTDS, it has several limitations. First, the samples used for data collection in this current study were recruited from four different universities located in a single geographical region (the Black Sea region) of Turkey. Turkey has seven different geographical regions that may have diverse cultural characteristics. So, it can be said that the samples for this current study had limited diversity in terms of some characteristics such as parents' educational background, socio-economic status, and cultural background which are important key factors that can be related to CT (Ennis, 1987b; Kennedy et al., 1991; McPeck, 2016). Therefore, future studies can be carried out to investigate the psychometric properties of the TV-CTDS with more diverse samples recruited from different geographical regions of Turkey which will help us generalize the factorial structure of the TV-CTDS to other Turkish settings. Second, in this current study, the empirical validity of the TV-CTDS was not investigated which can be tested by calculating the correlations between the TV-CTDS and other antecedents or outcomes of CT such as decision making, rational thinking, problem solving, or academic achievement. Therefore, it is

advised that future studies can be conducted to investigate the relationship between the scores produced by the TV-CTDS and other several antecedents or outcomes of CT to provide evidence for the empirical validity of the TV-CTDS. Thirdly, although the one-factor model presented a worse data fit than the two-factor model in many previous validation studies (Akın et al., 2015; Gerdts-Andresen et al., 2022; Sosu, 2013; Thomas & Hayes, 2021), including this current study, there are also some other studies concluding the one-factor model of the CTDS produced a good data fit (Bravo et al., 2020; Luiz et al., 2021; Yockey, 2016). Therefore, the one-factor and two-factor models of the CTDS should be investigated comparatively while investigating the factor structure of the CTDS in future studies. Lastly, the TV-CTDS is an instrument designed to measure individuals' CTD. CT is not only about cognitive skills such as deduction, making inferences, and interpretation, but also about some dispositions or attitudes towards using these cognitive skills (Ennis, 1987a; Paul, 1990). Having only high CTS or CTD does not make the individual a good critical thinker and both CTS and CTD are important for individuals to be a good critical thinker (Pascarella & Terenzini, 1991). Therefore, it is advised to use the TV-CTDS in combination with other cognitive instruments designed to measure CTS to gain a better understanding of individuals' CT.

6. References

Akın, A., Hamedoğlu, M. A., Arslan, S., Akın, Ü., Çelik, E., Kaya, Ç., & Arslan, N. (2015). The adaptation and validation of the Turkish version of the Critical Thinking Disposition Scale (CTDS). *The International Journal of Educational Researchers, 6*(1), 31-35.

Beyer, B. (1995). *Critical thinking*. Phi Delta Kappa Educational Foundation.

Bialosiewicz, S., Murphy, K., & Berry, T. (2013). *An introduction to measurement invariance testing: Resource packet for participants*. Claremont Evaluation Center.

Bondy, K. N., Koenigse, I. A., Ishee, J. H., & Williams, B. G. (2001) Psychometric properties of the California critical thinking tests. *Journal of Nursing Measurement, 9,* 309-328.

Bravo, M. J., Galiana, L., Rodrigo, M. F., Navarro-Perez, J. J., & Oliver, A. (2020). An adaptation of the Critical Thinking Disposition Scale in Spanish youth. *Thinking Skills and Creativity, 38*(100748), 1-12.

Chan, D. (2011). Advances in analytical strategies. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (pp. 85–113). American Psychological Association.

Ennis, R. H. (1987a). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, *18*(3), 4-10.

Ennis, R. H. (1987b). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). Henry Holt & Co.

Ennis, R. H. (1991). Goals for a critical thinking curriculum and its assessment. A. L. Costa (Ed.). In *Developing minds: A resource book for teaching thinking*. Ascd.

Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell critical thinking tests Level X and Level Z manual*. The Critical Thinking Co.

Epstein, R. L., & Kernberger, C. (2012). *Critical thinking.* Advanced Reasoning Forum.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction - executive summary - the Delphi report*. The California Academic Press.

Facione, P. A., & Facione, N. (1992). *The California Critical Thinking Dispositions Inventory (CCTDI) and the CCTDI test manual*. California Academic Press.

Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement and relationship to critical thinking skill. *Journal of Informal Logic, 20*(1), 61-84.

Fisher, A. (2001). *Critical thinking: An introduction*. Cambridge University Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Gerdts-Andresen, T., Hansen, M. T., & Grondahl, V. A. (2022). Educational effectiveness: Validation of an instrument to measure students' critical thinking and disposition. *International Journal of Instruction, 15*(1), 685-700.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective.* Pearson Education.

Halpern, D. F. (2003). *Thought & knowledge: An introduction to critical thinking*. Lawrence Erlbaum Associates Publishers.

Huitt, W. (1998). *Critical thinking: An overview. Educational Psychology Interactive*. Valdosta State University.

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

Irani, T., Rudd, R., Gallo, M., Ricketts, J., Friedel, C., & Rhoades, E. (2007). *Critical thinking instrumentation manual.* http://step.ufl.edu/resources/critical_thinking/ctmanual.pdf.

Kakai, H. (2003). Re-examining the factor structure of the California Critical Thinking Disposition Inventory. *Perceptual and Motor Skills, 96*, 435-483.

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). Critical thinking: Literature review and needed research. In L. Idol & B.F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform* (pp. 11–40). Erlbaum & Associates.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. The Guilford Press.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice, 32*(3), 131-137.

Luiz, F. S., Leite, I. C. G., Carvalho, P. H. B., Püschel, V. A. A., Braga, L. M., Dutra, H. S., Sanhudo, N. F., & Carbogim, F. C. (2021). Validity evidence of the Critical Thinking Disposition Scale, Brazilian version. *ACTA Paulista de Enfermagem, 34*, 1-8.

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341.

McPeck, J. E. (2016). *Critical thinking and education*. Routledge.

Muthen, L. K., & Muthen, B. O. (2010). *Mplus user's guide* (6th ed.). Authors.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Orhan, A. (2022). California Critical Thinking Disposition Inventory: Reliability generalization meta-analysis. *Journal of Psychoeducational Assessment, 40*(2), 202–220. https://doi.org/10.1177/07342829211048962

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. Jossey-Bass.

Paul, R. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world*. Center for Critical Thinking and Moral Critique.

Profetto-McGrath, J., Hasketh, K. L., Lang, S., & Estabrooks, C. A. (2003). A study of critical thinking and research utilization among nurses. *Western Journal of Nursing Research, 25*(3), 322-337.

Profetto-McGrath, J. (2003). The relationship of critical thinking skills and critical thinking dispositions of baccalaureate nursing students. *Journal of Advanced Nursing, 43*(6), 569–577.

Ruggerio, V. R. (1990). *Beyond feelings: A guide to critical thinking*. Mayfield Publishing Company.

Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and education*. Routledge.

Sosu, E. M. (2013). The development and psychometric validation of a Critical Thinking Disposition Scale. *Thinking Skills and Creativity, 9*, 107-119.

Stassen, M.L. A., Herrington, A., & Henderson, L. (2011). Defining critical thinking in higher education: Determining assessment fit. *To Improve the Academy, 30*(1), 126-141.

Stedman, N., & Adams, B. (2012). Identifying faculty's knowledge of critical thinking concepts and perceptions of critical thinking instruction in higher education. *NACTA Journal, 56*(2), 9-14.

Thomas, C. L., & Hayes, A. R. (2021). Using exploratory structural equation modeling to investigate the construct validity of the Critical Thinking Disposition Scale. *Journal of Psychoeducational Assessment, 39*(5), 640–648.

Tican, C., & Taspinar, M., (2015). The effects of reflective thinking-based teaching activities on pre-service teachers' reflective thinking skills, critical thinking skills, democratic attitudes, and academic achievement. *Anthropologist, 20*(1, 2), 111-120.

Walsh, C. M., & Hardy, R. C. (1997). Factor structure stability of the California Critical Thinking Disposition Inventory across sex and various student majors. *Perceptual and Motor Skills, 85*, 1211-1228.

Walsh, C. M., Seldomridge, L. A., & Badros, K. K. (2007). *Perceptual and Motor Skills, 104*, 141-151.

Watson, G., & Glaser, M. E. (1994). *Watson-Glaser critical thinking appraisal form S manual*. The Psychological Corporation.

Yockey, R. D. (2016). Validation study of the Critical Thinking Dispositions Scale: A brief report. *North American Journal of Psychology, 18*(1), 101–106.