Validation of Critical Thinking Test on Slovak Conditions

Erika Sustekova [1*], Milan Kubiatko [2], Muhammet Usak [3]

[1] University of Zilina, Faculty of Humanities, Department of Mediamatics and Cultural Heritage, Univerzitna 1, 010 26 Zilina, SLOVAKIA
[2] J. E. Purkyne University, Faculty of Science, Department of Biology, Ceske mladeze 8, 400 96 Usti nad Labem, CZECH REPUBLIC
[3] Kazan Federal University, Institute of Pedagogy and Psychology, Kazan, RUSSIA

ABSTRACT
Critical Thinking is a generally recognized educational ideal at all levels of the educational process. This study validated the critical thinking test on Slovak conditions. Data were collected from 50 respondents studying at university. Bachelor's and Master's students from all grades, aged 21 to 36 (x = 23.00; SD = 2.84) were represented. Model fits, normality, reliability and exploratory factor analysis were used. The results showed that the critical thinking test is valid and reliable. The implications of the test for next using is discussed.

Keywords: critical thinking test, instrument development, model fit, university students

INTRODUCTION

Critical Thinking (henceforth CT) is a generally recognized educational ideal at all levels of the educational process. It is considered one of the most important skills and dispositions not only in the sphere of tertiary education, but is also an important skill in everyday life situations, in the professional life of the individual – it is increasingly required not only in top management, but also in a wide range of professions and last but not least and the new media brought about by the 21st century.

Many experts also consider it a liberating force in education, a precondition for maintaining the ideals of democratic society, a source of civic engagement, decision-making and the ability to respond to rapidly changing world around (Ennis, 1996; Facione, 2011; Halpern, 1998, Paul & Elder, 2014), which corresponds to the mission of many universities in the world. It is also embedded in the international frameworks for the competences of the 21st century (Voogt & Roblin, 2012).

However, with the increasing importance of critical thinking, the urgency of its adequate assessment is also on increase. While there are several validated tests for measuring critical thinking disposition, the vast majority of them have roots in the USA, with the language mutations of these tests which are either not existing or not sufficiently validated within the country of use. Studies from various countries around the world (Ahdika, 2017; Basha, Drane, & Light, 2016; Cankaya & Serin, 2018; Chikiwa & Schäfer, 2018; de Bie, Wilhelm, & van der Meij, 2015; Inel Ekici, 2017; Ioannou & Malafantis, 2019; Liu et al., 2016; Miller & Olthouse, 2013; Rodrigues Franco, Soares Costa, & da Silva Almeida, 2018; Sendag, Erol, Sezgin, & Dulkadir, 2015; Wartono, Hudha, & Batlolona, 2018; Yu, Lin, & Chang, 2017) discuss the importance of this issue.

In connection with this, we will standardize and adapt the *Critical Thinking Prerequisites Test* for the conditions of the Slovak setting. The test was designed as a requirement arising from the absence of a similar instrument in the Czech Republic (where the situation is similar to that in the Slovak Republic). This test was chosen because of the linguistic and cultural closeness of both countries and was piloted on a sample of students n = 203. This study focuses on translation, adaptation and determination of psychometric properties of the *Critical Thinking Prerequisites Test* for the requirements of the University of Žilina in the Slovak Republic.

Contribution of this paper to the literature

- The paper illustrates the importance of critical thinking among university students.
- The model fit indices for the factor analysis was reasonably good.
- The test of critical thinking was adapted on the conditions of Slovak Republic.

LITERATURE REVIEW

Despite the apparent popularity of critical thinking (CT), its clear definition remains even today, more than 100 years after John Dewey (1910) introduced the concept of reflexive thinking (from which many other definitions of CT were subsequently derived), ambiguous and problematic. Since CT is a multidimensional concept and experts from different areas perceive it primarily through the lens of their operations, this task poses many complications. Bensley (2011, p. 2) states that "*many view CT as a legitimate construct, it remains a loose conglomeration of concepts without a clear claim to scientific legitimacy. This has impeded progress in the scientific study of CT, in general, and in the determination of what are the most effective ways to teach and assess it, in particular*". However, we cannot fully agree with this idea - despite the declared multilateralism, the definitions basically correspond to each other. Hitchcock (2017) who points to many common aspects within the definition of CT at the same time adds that "we should not be surprised by the apparent absence of consensus. New domains are normally the subject of numerous definitions before a broad consensus is reached". Similarly, Paul and Elder (2007) refer to the issue of defining the CT concept, claiming that the concept of critical thinking can be expressed in different definitions, depending on purpose; however, its essence is always the same.

Research conducted by a group of experts from various fields attempted to bring research consensus into a considerable degree of ambivalence in this regard. The experts participated in the interactive panel of experts (the method is known as "the Delphi method") between 1988 and 1989 and the result was a paper by Peter Facione (1990) *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction (The Delphi Report)*. On the basis of this paper, experts have defined CT very comprehensively as "*purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based*" (Facione, 1990, p. 3). Despite the existence of many diverse concepts of critical thinking (e.g. Brookfield, 1987; Ennis, 1985; Facione, 1990; Halpern, 1998; McPeck, 1981; Mulnix, 2012; Paul, 1989; Paul & Elder, 2007; Scriven & Paul, 2008; Stenberg, 1985) on the basis of an extensive theoretical research, it is possible to claim that their fundamentals remain overwhelmingly consistent and the definition of the Delphi Report sufficiently covers the background of CT.

With an increasing importance and interest in critical thinking, new ways to measure and evaluate this ability have also emerged (Butler, 2012; Halpern, 2010). For this reason, there are a number of divergent methods that are used for this purpose, while the most common instrument is psychometric testing and at present, we commonly encounter them in the postgraduate, professional, or management sectors. The vast majority of critical thinking tests take the multiple-choice selection format, covering a wide range of critical thinking aspects, with a time limit being predominantly designed. The majority of the critical thinking assessments that are used originate in the United States as is stated in the introduction to the study.

McMillan (1987) investigated the CT studies and concluded that the most commonly used assessment tool is Watson-Glaser Critical Thinking (WGCT) - the standard version consists of multiple-choice questions for a test that takes an hour. There are two versions (A and B) that are supposed to be equivalent and can be used to measure changes in critical thinking over a period of time (Glaser & Watson, 2010). The California Critical Thinking Skills Test: College Level (Facione 1990) is often used by U.S. academic assessments. This critical thinking test is aimed at university students and it evaluates analysis, inference, evaluation, deductive and inductive thinking with the help of multiple-choice items. Another assessment tool that is used is Cornell Critical Thinking Tests (Ennis, Millman, & Tomko, 2005), which also features multiple choice questions and two rating levels: Level X is aimed at secondary school students and it evaluates induction, deduction, credibility, and identification of assumptions; Level Z is aimed at university students and adults and focuses on assessing induction, deduction, credibility, and identification of assumptions, errors, definitions, predictions, and experimental planning. The Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985) is a test based on open questions which is focused on assessing general skills of argumentation in CT.

One of the most recent and commonly used assessments was The Halpern Critical Thinking Assessment - HCTA (Halpern, 2012), which is considered a turning point in CT's assessment (Franco, Costa, & Almeida, 2018), mainly due to the format combining multiple-choice items, which enables a more comprehensive assessment of CT, the content aspect focusing on the real-life scenario and last but not least it covers not only CT skills but also

dispositions. The HCTA test evaluates five dimensions that correspond to the five skills presented by Halpern's theoretical CT model: verbal reasoning; argument analysis; thinking as hypothesis testing; probability and uncertainty; decision making and problem solving. At the same time, Halpern notes that it is culturally neutral and therefore, it can be used across domains and states (Halpern, 2012).

In spite of the wide range of critical thinking skills assessment, however, there are a number of shortcomings in these assessments which are presented by Liu et al. (2016), describing in particular the lack of initial conceptualization or non-equivalent difficulty of the forms in tests involving multiple forms. Also, as one of the problems, the study mentions the use of exclusively discrete items with multiple selection options, which may not be able to capture the more sophisticated CT skills and the higher cognitive processes required for CT, which are typically reflected in extended item types. In addition, close - ended answers and test forms based on multiple - choice are criticized by other authors (Ennis, 1996; McPeck, 1984; Nicol, 2007; Popham, 2003), however, on the other hand, it is undeniable that the reliability of this form of testing is high, whether by eliminating the evaluator's subjectivity and also owing to the same conditions for all tested persons (Benjamin, 2014). On this basis, Liu et al. (2016) specified the need for a clear operational definition and the theoretical basis of assessment in the development of a new assessment tool, focusing on the most important components in the critical thinking frameworks and ensuring technical validity through the relevant psychometric aspects of evaluation. Responding to the need for the evaluation of critical thinking across the global learning process, several new studies have originated (Liu et al., 2016; Lopez, 2012) and they deal with the adaptation of the already existing tests to a specific linguistic and cultural environment (Basha, Drane, & Light, 2016; de Bie Wilhelm & van der Meij, 2015; Rodrigues Franco, Soares Costa, & da Silva Almeida, 2018; Verburgh et al., 2013).

Lopez (2012) has developed and validated a critical thinking test for tertiary Philippine students entitled CEU - Lopez Critical Thinking Test (2012), which consists of 87 items and is considered a multi-aspect test of general critical thinking knowledge (Lopez & Asilo, 2014).

The psychometric quality of the Dutch language version of the Cornell Critical Thinking Test (CCTT) and Halpern's Critical Thinking (HCTA) was examined in a study by Verburgh et al. (2013) on a sample of Dutch-speaking university students studying pedagogical science. The results showed a higher validity of the content and preferences of students for HCTA, but none of the tests showed high overall reliability. De Bie, Wilhelm and van der Meij (2015) in their study also focused on the psychometric features of the Dutch version of the Halpern Critical Thinking Assessment (HCTA). The study involved students of communication and psychology (N = 240). In this particular case, the reliability of HCTA has proven to be adequate. At the same time, respondents had the opportunity to provide feedback after the end of the evaluation, which was used by 55 respondents, 37 of whom stated that the complex and incorrect syntax and the use of scientific language made it difficult to understand certain HCTA scenarios and issues. Accordingly, the researchers concluded that the issues that were raised could have affected the clarity of certain items. A lower average compared to the U.S. standard could also be due to translation and cultural differences and these aspects could affect test scores.

A study on translation, adaptation and assessment of the psychometric properties of HCTA was also conducted in Portugal (Rodrigues Franco, Soares Costa, & da Silva Almeida, 2018), with a sample of 333 university students. The aim was also to analyze whether there were differences in overall score by disciplinary and academic level. The obtained results supported the validity of the Portuguese version of HCTA, and these findings are of interest not only for Portugal but also in the context of multicultural testing of HCTA. So far, there are no validated critical thinking tests in Arabic. In response to this situation, Basha, Drane, and Light (2016) focused on assessing the suitability of the American Critical -Thinking Assessment Test (CAT) for the use in Palestine. The test was piloted with university students in English (n = 30) and four questions were piloted in Arabic (n = 48). Students responded favorably and the scores were comparable to those in the USA. Only two students found that content is problematic and the CAT assessment has the potential to be a tool for assessing critical thinking in Palestinian higher education, and an extensive CAT test in Arabic should be conducted.

METHODOLOGY

The Research Sample

50 respondents studying at university participated in the research investigation. Bachelor's and Master's students from all grades, aged 21 to 36 (x = 23.00; SD = 2.84) were represented. At the time of research procedure, the respondents studied at the Faculty of Humanities of the University of Zilina. Respondents were selected by the available selection and the sample size is sufficient for the process of conducting research investigations (Newby, 2014).

Research Tool

In order to achieve the set goal, a quantitative approach to data acquisition and analysis through a critical thinking test as a research technique was adopted. Before the research tool was administered to students, research was conducted when researchers involved in methodology were asked to evaluate the critical thinking test. After the realization of the pre-research procedure, certain items were modified with regards to their stylistic character. These modifications did not change the content of the edited items. The original research tool has already been employed in a final thesis and the author of the thesis has allowed the use of a critical thinking test at discretion.

Research Tool Administration

Tests were distributed to students through an electronic link provided via email or as a social networking link. Students were assured of anonymity and were also familiarized with the purposes of using the obtained data. The completion time did not exceed 30 minutes and all tests were completed in such a way that they could be included in research process.

Data Analysis

After obtaining the tests, the data were transferred to MS Excel. Subsequently, the reliability of the data was determined by means of the Cronbach alpha (α) coefficient. The determination of mutual correlation between individual items was identified as well. The explorative factor analysis was employed with the aim to determine the construct validity. Also, there were used Chi Square/df ratio, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Residual (SRMR) as the indicators for model fit. In determining the psychometric characteristics of the research tool, the following source was followed - Cockley et al. (2001) as well as other research articles addressing the psychometric properties of the research tool (e.g. Osteras, Sigmundsson, & Haga, 2018). The data were processed in SPSS 22.0, STATISTICA 10.0, and in Microsoft Excel 10.0.

RESULTS

The Item Analysis and the Determination of Normality

In the initial stage, the item analysis was performed for all 40 items of the Slovak critical thinking test. The results are presented in **Table 1**, which shows that a maximum range of 0 to 1 was found. The average score was the lowest for item No. 6 and the highest for item No. 30.

Table 1. The Item Analysis of the Slovak version of Critical Thinking Test

Question Number	Average	Standard Deviation	Minimum	Maximum
1	0.81	0.38	0.00	1.00
2	0.63	0.48	0.00	1.00
3	0.65	0.47	0.00	1.00
4	0.79	0.40	0.00	1.00
5	0.33	0.46	0.00	1.00
6	0.19	0.38	0.00	1.00
7	0.39	0.48	0.00	1.00
8	0.28	0.44	0.00	1.00
9	0.56	0.49	0.00	1.00
10	0.24	0.42	0.00	1.00
11	0.35	0.47	0.00	1.00
12	0.75	0.43	0.00	1.00
13	0.37	0.48	0.00	1.00
14	0.47	0.49	0.00	1.00
15	0.40	0.48	0.00	1.00
16	0.37	0.48	0.00	1.00
17	0.51	0.49	0.00	1.00
18	0.45	0.49	0.00	1.00
19	0.42	0.49	0.00	1.00
20	0.47	0.49	0.00	1.00
21	0.38	0.48	0.00	1.00
22	0.47	0.49	0.00	1.00
23	0.28	0.44	0.00	1.00
24	0.31	0.46	0.00	1.00
25	0.44	0.49	0.00	1.00
26	0.54	0.50	0.00	1.00
27	0.56	0.49	0.00	1.00
28	0.47	0.49	0.00	1.00
29	0.45	0.49	0.00	1.00
30	0.88	0.32	0.00	1.00
31	0.38	0.48	0.00	1.00
32	0.47	0.49	0.00	1.00
33	0.33	0.46	0.00	1.00
34	0.56	0.49	0.00	1.00
35	0.35	0.47	0.00	1.00
36	0.37	0.48	0.00	1.00
37	0.58	0.49	0.00	1.00
38	0.58	0.49	0.00	1.00
39	0.63	0.48	0.00	1.00
40	0.45	0.49	0.00	1.00

The values presented in the table indicated the distribution of data approximately in accordance to the Gaussian curve. The assumption was confirmed by the normality test conducted in STATISTICA 10.0. The result of the test was that at the commonly used significance level $p < 0.05$, the data were very likely to be from a standard distribution. The Shapiro-Wilks test value was $W = 0.95$ ($p = 0.06$). The data distribution is presented in **Figure 1**. The above mentioned facts pointed to another possible use of the test by means of parametric statistical methods on condition of standard data distribution.
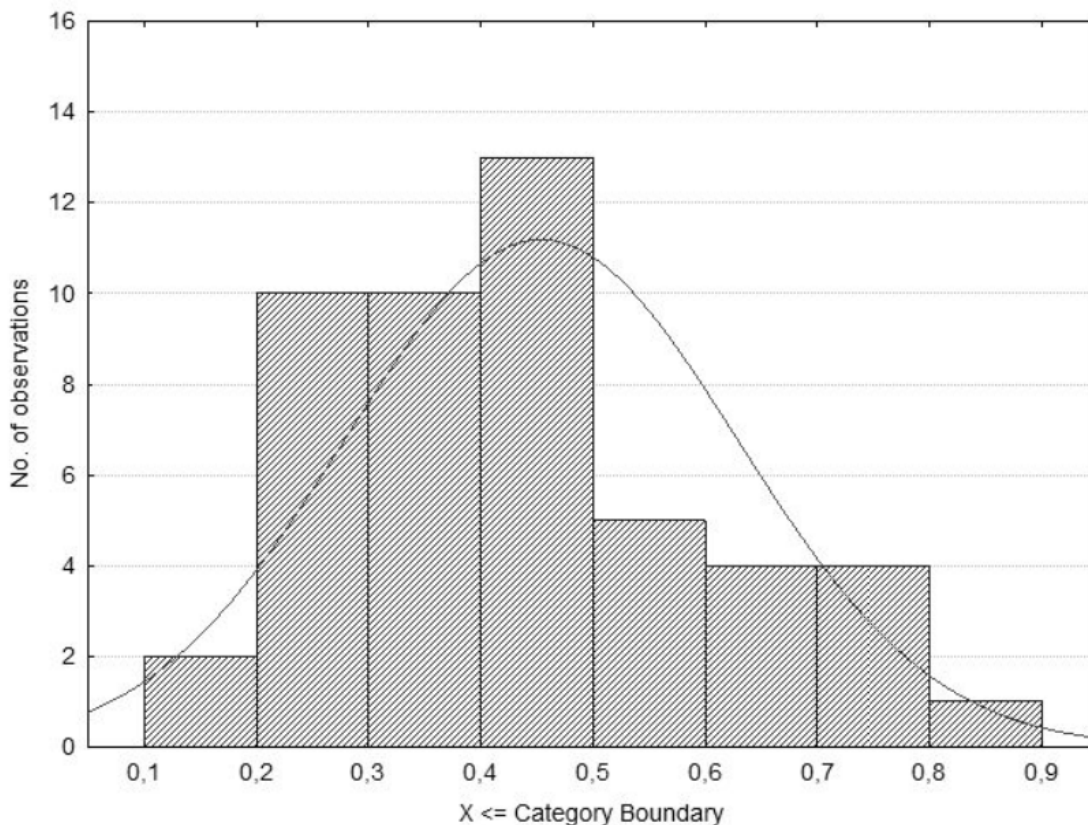
Figure 1. Data Distribution

Table 2. Model fit indices

Fit indices	Values
X^2/ df	2.88
CFI	0.92
TLI	0.94
RMSEA	0.07
SRMR	0.07

CFI – Comparative Fit Index; TLI – Tucker-Lewis index; RMSEA – Root Mean Square Error of Approximation; SRMR – Standardized Root Mean Square Residual

Reliability

In order to determine the reliability of the research tool, two techniques were used. The first was the Cronbach alpha (α) coefficient, the value of which was α = 0.83, which indicates the overall internal consistency of the questionnaire relative to the number of items and respondents. The values of the α coefficient were also determined for the individual factors listed in the following section on factor analysis.

The second technique was the KR21 (Kuder-Richardson21) coefficient, which was used for the entire test where it was 0.82, and for each dimension. For the analytical part, the value was 0.77, for the evaluation 0.70 and for the inference 0.79, which also pointed to the reliability of the research tool.

Fit Model

Fit indices for the EFA are summarized in the **Table 2**. Chi Square/df ratios was lower than the suggested ratio (3.00). CFI and TLI were higher than the criterion of 0.90. RMSEA value was lower than 0.08, which fell within the reasonable fit of 0.05 – 0.08. SRMR value was lower than 0.08. To sum up, the model fit indices was reasonably good.

Table 3. Results of the factor analysis

Item Number	α	I	II	III
I. The Analytical Part	**0.62**			
1		**0.49**	0.17	-0.03
3		**0.48**	0.06	0.14
4		**0.52**	0.16	0.11
11		**0.41**	0.29	0.01
14		**0.56**	0.37	0.03
15		**0.46**	-0.05	0.06
16		**0.44**	0.25	0.07
20		**0.46**	0.06	-0.14
24		**0.56**	0.08	-0.07
34		**0.49**	0.13	-0.02
36		**0.66**	0.10	-0.08
40		**0.55**	0.01	0.20
II. The Evaluative Part	**0.71**			
7		0.18	**0.42**	0.11
8		0.19	**0.68**	-0.28
9		-0.14	**0.45**	-0.13
10		-0.05	**0.44**	0.04
17		-0.18	**0.51**	-0.29
19		0.20	**0.49**	0.31
21		-0.18	**0.48**	0.20
22		0.02	**0.49**	0.10
23		0.24	**0.61**	-0.04
25		0.01	**0.67**	-0.05
26		-0.05	**0.52**	0.21
27		0.01	**0.41**	-0.02
28		0.15	**0.48**	-0.06
37		0.17	**0.49**	-0.01
38		0.28	**0.40**	0.21
39		-0.01	**0.44**	0.24
III. The Inferential Part	**0.63**			
2		0.39	0.15	**0.43**
5		0.14	-0.20	**0.59**
6		-0.08	-0.06	**0.71**
12		-0.30	0.31	**0.52**
13		0.20	0.28	**0.49**
18		0.19	0.22	**0.48**
29		0.08	-0.12	**0.44**
30		0.31	-0.09	**0.58**
31		0.21	-0.08	**0.59**
32		0.37	-0.15	**0.43**
33		-0.26	0.23	**0.66**
35		-0.06	0.05	**0.49**
Figure		6.17	2.78	2.57
Variance Percentage		37.44	16.95	6.42

Factor Analysis

The Explorative Factor Analysis (EFA) was used in order to validate the construct validity, as research to verify the presented research tool has not been published up to this point, its only source was a final thesis. The suitability of using factor analysis on the obtained data was verified by the Kaiser-Meyer-Olkin (KMO) test and the Bartlett sphericity test. Both tests used values to use the factor analysis. The Bartlett test value was $\chi^2 = 3758.22$ (p <0.001) and the KMO test was 0.81. The results of the factor analysis are presented in **Table 3**, together with the values of the Cronbach alpha (α) coefficient, for the individual parts of the critical thinking test. The values indicate a high reliability of the research tool.

DISCUSSION AND CONCLUSION

The aim of the study was to perform a psychometric verification of the Slovak version of the critical thinking test among university students. The research tool was created in the Czech Republic and was used only for the purpose of a final thesis; therefore, it has not been discussed in any published research. The usual procedures were followed with the aim to determine the psychometric properties.

With regards to the internal consistency of the Slovak version of the critical thinking test expressed by the Cronbach alpha coefficient of 0.83, it can be stated that the test shows a high reliability. Similarly, reliability testing was also performed using the Kuder Richardson KR21 test. This test was primarily used to determine the reliability of research tools that featured dichotomous coding questions (Tucker, 1949). The values that were found showed the reliability of the critical thinking test (Charter, 2007). A similar statement can also be made on the basis of the alpha values that determine the reliability of the research tool; George and Mallery (2003) presented the value of alpha above 0.80 as optimal. It is also possible to discuss the reliability of a research tool when the values found are related to the value of 0.70 reported by Nunnally (1977) as sufficient to consider the research tool to be reliable. The Cronbach alpha coefficient compared to the KR21 coefficient is more accurate and reliable; it can determine more accurately the reliability, even with a research tool that produces dichotomous data. Naturally, even the Cronbach alpha coefficient does not belong to error-free tests, it has its shortcomings, but it can measure the reliability exactly as is stated e.g., in Sijtsma (2009). When evaluating the *s* individual factors, the alpha values for individual factors ranged from 0.62 - 0.71, the stated values declare the reliability of the individual factors. As is claimed by Ferketich (1991) and Kline (1993), when classifying the items in the questionnaire into respective dimensions, it is expected that the alpha value of the dimensions will decrease, while the reliability limit of the individual dimensions is determined as $\alpha = 0.50$.

Before EFA the fit indices were used due the testing accuracy of research tool. There were used chi square / df ratio. This index should have value lower than 3.00 (Carmines & McIver, 1983), in the case of our data, the value was lower than suggested value (2.88). CFI and TLI indices should have values higher than 0.90 (Bentler, 1990). In our case in the both cases values were higher than suggested (0.92 resp. 0.94). RMSEA fit index recognises that models can only ever be approximately correct. If the value of RMSEA is lower than 0.08, it suggested reasonable model fit (Chen et al., 2008). In our case the value was 0.07. SRMR (The standardized root mean-square residual) is an absolute measure of fit is defined as the standardized difference between the observed correlation and the predicted correlation (Joreskog & Sorborn, 1981). As in previous case, if the value of SRMR is lower than 0.08, it suggested reasonable model fit. In our case the value was 0.07.

The exploratory factor analysis was used for the determination of the construct validity and for the determination of the classification of items into dimensions/ factors. This analysis classified the items into three factors. In the process of research, the conditions for the application of factor analysis – the calculation of the KMO test and Bartlett's sphericity test were followed. The percentage of variance also indicates the correct use of EFA. As a percentage of variance, the first dimension should account for at least 20% of the total variance and the difference between the second and third dimension must be less than the difference between the first and second factor (Reckase, 1979). The dimensions of our research investigation explained 60.81% of the total variance and the Reckase rule was also followed.

The limits of the present study include the scope of the research sample obtained from the ordinary population of secondary school pupils in Slovakia. The total number of respondents is lower than it should be, also considering the number of items in the research tool as well as the use of factor analysis. The issue which is related to the number of respondents eligible for factor analysis is discussed by Field (2000) who states that it is 10 to 15 respondents per item in a research tool that must be involved in a research procedure in order to apply the factor analysis. On the basis of this fact, it is recommended to use the data obtained exclusively in a given research context. In order to fully standardize the critical thinking test for Slovak conditions, particular attention needs to be paid to the re-verification of another research sample and also to the use of additional criteria for the determination of the validity of the research tool.

REFERENCES

Ahdika, A. (2017). Improvement of Quality, Interest, Critical, and Analytical Thinking Ability of Students through the Application of Research Based Learning (RBL) in Introduction to Stochastic Processes Subject. *International Electronic Journal of Mathematics Education, 12*(2), 167-191.

Basha, S., Drane, D. L., & Light, G. (2016). Adapting the Critical Thinking Assessment Test for Palestinian Universities. *Journal of Education and Learning, 5*(2), 60-72. https://doi.org/10.5539/jel.v5n2p60

Benjamin, R. (2014). Two Questions About Critical-thinking Tests in Higher Education. *Change: The Magazine of Higher Learning, 46*(2), 24–31. https://doi.org/10.1080/00091383.2014.897179

Bensley, D. A. (2011). Rules for reasoning revisited : toward a scientific conception of critical thinking. In: *Critical thinking*. Hauppauge, NY: Nova Science Publishers.

Bentler, P. M. (1990). Comparative Fit Indices in Structural Models. *Psychological Bulletin, 107*, 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Brookfield, S. (1987). *Developing Critical Thinkers. Challenging Adults to Explore Alternative Ways of Thinking and Acting*. San Francisco: Jossey-Bass Publishers.

Butler, H. A. (2012). Halpern Critical Thinking Assessment Predicts Real-World Outcomes of Critical Thinking. *Applied Cognitive Psychology, 26*, 721-729. https://doi.org/10.1002/acp.2851

Cankaya, S., & Serin, O. (2018). Analysis of the Relationship between School Heads' Leadership, Problem Solving and Critical Thinking Skills. *Online Journal of Communication and Media Technologies, 8*(1), 26-40. https://doi.org/10.29333/ojcmt/2361

Carmines, G. E., & McIver, J. P. (1983). An Introduction to the Analysis of Models with Unobserved Variables. *Special Issue on Modeling, 9*(1), 51-102.

Chen, F., et al. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points inRMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research, 36*(4), 462-494. https://doi.org/10.1177/0049124108314720

Chikiwa, C., & Schäfer, M. (2018). Promoting Critical Thinking in Multilingual Mathematics Classes through Questioning. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(8), em1562. https://doi.org/10.29333/ejmste/91832

de Bie, H., Wilhelm, P. & van der Meij, H. (2015). The Halpern Critical Thinking Assessment: Towards a Dutch Appraisal of Critical Thinking. *Thinking Skills and Creativity, 17*, 33-44. https://doi.org/10.1016/j.tsc.2015.04.001

Dewey, J. (1910). *How We Think*. Boston: D.C. Heath & Co. https://doi.org/10.1037/10903-000

Ennis, R. H. (1985). Critical Thinking and the Curriculum. *National Forum: Phi Kappa Phi Journal, 65*(1), 28-31.

Ennis, R. H. (1996). *Critical Thinking*. Upper Saddle River: Prentice Hall.

Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test [Measurement instrument]*. Pacific Grove, CA: Midwest Publications.

Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell Critical Thinking Tests: Administration Manual* (5th Ed.). Seaside, CA: The Critical Thinking Co. (formerly Midwest Publications).

Facione, P. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction (The Delphi Report)*. Educational Resources Information Center (ERIC).

Facione, P. A. (2011). *Critical thinking: What it is and why it counts*. Millbrae, CA: Measured Reasons and The California Academic Press.

Halpern, D. (1998). Teaching Critical Thinking for Transfer Across Domains. Dispositions, Skills, Structure Training, and Metacognitive Monitoring. *American Psychologist. USA: American Psychological Association, 53*(4), 449-455. https://doi.org/10.1037/0003-066X.53.4.449

Halpern, D. F. (1998). Teaching Critical Thinking for Transfer Across Domains: Disposition, Skills, Structure Training, and Metacognitive Monitoring. *American Psychologist, 53*(4), 449-455. https://doi.org/10.1037/0003-066X.53.4.449

Halpern, D. F. (2012). *Halpern Critical Thinking Assessment [Measurement in-strument]*. Mödling: Schuhfried. https://doi.org/10.1037/t10940-000

Hitchock, D. (2017). *On Reasoning and Argumen*. Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-53562-3

Inel Ekici, D. (2017). The Effects of Online Communities of Practice on Pre-Service Teachers' Critical Thinking Dispositions. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(7), 3801-3827. https://doi.org/10.12973/eurasia.2017.00759a

Ioannou, A. X., & Malafantis, K. D. (2019). The Contribution of the Literary Works of François Fénelon 'The Adventures of Telemachus, Son of Ulysses' and of Voltaire 'Candide', to the Cultivation of Humanitarian Awareness. *Pedagogical Research, 4*(4), em0042. https://doi.org/10.29333/pr/5879

Joreskog, K., & Sorborn, D. (1981). *LISREL V: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*. Chicago: National Educational Resources.

Liu O. L., et al. (2016). Assessing Critical Thinking in Higher Education: the HEIghten™ Approach and Preliminary Validity Evidence. *Assessment & Evaluation in Higher Education, 41*(5), 677-694. https://doi.org/10.1080/02602938.2016.1168358

Lopez, M. Y. (2012). *The CEU-Lopez Critical Thinking Test Manual*. Research and Evaluation Office, Centro Escolar University, Manila.

Lopez, M. Y., & Asilo, M. V. (2014). Development and Validation of The CEU-Lopez Critical Thinking Test. *Inquiry: Critical Thinking Across the Disciplines, 29*(1), 32-55. https://doi.org/10.5840/inquiryct20142914

McMillan, J. H. (1987). Enhancing College Student Critical Thinking: A Review of Studies. *Research of Higher Education, 26*(1), 3-29. https://doi.org/10.1007/BF00991931

McPeck, J. (1984). The Evaluation of Critical Thinking Programs: Dangers and Dogmas. *Informal Logic, 6*(2), 9-13. https://doi.org/10.22329/il.v6i2.2727

McPeck, J. E. (1981). *Critical Thinking and Education*. New York: St. Martin's Press.

Miller, M. T., & Olthouse, J. (2013). Critical Thinking in Gifted Childrens' Offline and Online Peer Feedback. *Contemporary Educational Technology, 4*(1), 66-80.

Mulnix, J. W. (2012). Thinking Critically about Critical Thinking. *Educational Philosophy and Theory, 44*, 464-479. https://doi.org/10.1111/j.1469-5812.2010.00673.x

Nicol, D. (2007). E-assessment by Design: Using Multiple-choice Test to Good Effect. *Journal for Further and Higher Education, 31*, 53–64. https://doi.org/10.1080/03098770601167922

Østerås, B., Sigmundsson, H., & Haga, M. (2018). Psychometric Properties of the Perceived Stress Questionnaire (PSQ) in 15-16 Years Old Norwegian Adolescents. *Frontiers in psychology, 9*, 1850. https://doi.org/10.3389/fpsyg.2018.01850

Paul, R. W. (1989). Critical Thinking in North America: A New Theory of Knowledge, Learning, and Literacy. *Argumentation, 3*(2), 197–235. https://doi.org/10.1007/BF00128149

Paul, R., & Elder, L. (2007). *A Guide for Educators to Critical Thinking Competency Standard*. Tomales: The Foundation for Critical Thinking.

Paul, R., & Elder, L. (2014). *Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life* (2nd Ed.). Indianapolis. IN: FT Press.

Popham, W. J. (2003). *Test Better, Teach Better. The Instructional Role of Assessment*. Alexandria, VA: ASCD

Rodrigues Franco, A., Soares Costa, P., & da Silva Almeida, L. (2018). Translation, Adaptation, and Validation of the Halpern Critical Thinking Assessment to Portugal: Effect of Disciplinary Area and Academic Level on Critical Thinking. *Anales de Psicología / Annals of Psychology, 34*(2), 292-298. https://doi.org/10.6018/analesps.34.2.272401

Scriven, M., & Paul, R. (2007). *Defining Critical Thinking. Tomales: The Foundation for Critical Thinking*. Retrieved from http://www.criticalthinking.org/aboutCT/define_critical_thinking.cfm

Sendag, S., Erol, O., Sezgin, S., & Dulkadir, N. (2015). Preservice teachers' critical thinking dispositions and web 2.0 competencies. *Contemporary Educational Technology, 6*(3), 172-187.

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika, 74*(1), 107-120. https://doi.org/10.1007/s11336-008-9101-0

Sternberg, R. J. (1985). *Beyond IQ: a Triarchic Theory of Human Intelligence*. New York: Cambridge University Press.

Verburgh, A., et al. (2013). The Assessment of Critical Thinking Critically Assessed in Higher Education: A Validation Study of the CCTT and the HCTA. *Education Research International*, 1-13. https://doi.org/10.1155/2013/198920

Wartono, W., Hudha, M. N., & Batlolona, J. R. (2018). How Are The Physics Critical Thinking Skills of The Students Taught by Using Inquiry-Discovery Through Empirical and Theorethical Overview?. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(2), 691-697. https://doi.org/10.12973/ejmste/80632

Watson, G., & Glaser, E.M. (1991). *Watson-Glaser Critical Thinking Appraisal Manual*. London: Psychological Corporation.

Yu, K. C., Lin, K. Y., & Chang, S. F. (2017). The Development and Validation of a Mechanical Critical Thinking Scale for High School Students. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(5), 1361-1376. https://doi.org/10.12973/eurasia.2017.00675a

http://www.ejmste.com