İSTANBUL UNIVERSITY
CERRAHPAŞA

HAYEF: Journal of Education, 19(1); 9-17

DOI: 10.5152/hayef.2021.21031

Received: April 25, 2021
Accepted: September 22, 2021

HAYEF: Journal of Education

RESEARCH ARTICLE

Developing İstanbul Critical Thinking Scale

Burcu AYKAÇ, Ayşe Esra ASLAN

İstanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education, İstanbul, Turkey

Abstract

To conduct effective and efficient studies on critical thinking, this cognitive ability must be measured within the framework of certain standards. Based on this need, the development of the "Istanbul Critical Thinking Scale", which was based on the original Delphi report method, constitutes the main purpose of this study. The research was organized under psychological scale development methods. The population was from Istanbul University and Istanbul University-Cerrahpaşa. The sample ($N = 1031$) was randomly selected from the departments of these universities that accept students with different score types. As a result of the analysis, the item pool of five dimensions was prepared: interpretation, analysis, evaluation, inference, and explanation, and the "Istanbul Critical Thinking Scale" consisting of three dimensions (interpretation, evaluation, and inference) and a total of 25 items were developed. Analyses were carried out with confirmatory factor analysis. Diagonally weighted least squares were used as the estimation method. Some items were gradually removed from the scale by examining the standardized path coefficients and fit indices obtained as a result of the analyses. The reliability of the total and sub-dimensions of the scale was calculated by the stratified Cronbach alpha method. The fit indices of the scale were found to be at an acceptable level, and the internal consistency was 0.68.

Keywords: Critical thinking scale, critical thinking skills, scale development, university students

Introduction

Critical thinking is one of the most significant issues in education (Lai, 2011). Critical thinking has an important place not only in school life but also in situations that require making decisions in business life, social, and interpersonal contexts (deLantante, 2019b; Ennis, 1993; Erwin & Sebrel, 2003; Ku, 2009; Schneider & Liu, 2019). Thinking is a necessary tool of inquiry. Therefore, critical thinking has a liberating role in education and is a powerful resource in an individual's personal and civic life (Facione, 1990a).

It is one of the basic skills necessary to keep up with the times, especially in today's world where knowledge is increasing rapidly, and more importantly, the paradigms of dealing with knowledge are changing. Individuals must have the skills and willingness to think critically to solve problems constructively, draw logical conclusions, and make informed decisions. Critical thinking helps individuals to be more adaptable and flexible to cope with the constantly developing knowledge (Dwyer et al., 2014).

Critical thinkers have the competence to form their own opinions by asking the right questions, so they are not easily manipulated by others. This situation brings about being aware of their own needs and being able to make rational choices (Diker-Coşkun, 2018). For this reason, it is significant to carry out studies that will develop critical thinking so that individuals can make effective decisions both in their private lives and in issues that concern society. To conduct effective studies on critical thinking, this cognitive ability must be measured within the framework of certain standards. It is emphasized in the literature that since critical thinking has a multidimensional and complex structure, it is important to choose a comprehensive definition to increase the efficiency of the studies (Erwin & Sebrell, 2003; Facione, 1990a; Kline, 2005). Therefore, developing a measurement tool that will make up for the deficiencies of existing measurement tools based on an accurate and inclusive definition is essential.

In many educational reports, critical thinking is accepted as a key skill that should be developed and measured in higher education (American Association of Colleges and Universities, 2005; Higher Education Quality Council, Quality Enhancement Group, 1996; National Postsecondary Education Cooperative, 2000), but there are few studies on this issue in Turkey. To be able to conduct critical thinking studies on university students, it is necessary to develop nationally valid and reliable measurement tools. For this reason, this study aims to develop a critical thinking scale that is suitable for Turkey's culture to examine the critical thinking levels of university students.

Critical Thinking

Even though critical thinking has been an area of interest since ancient Greek philosophers, it has not found its rightful place in practice for a long time. At present, however, critical thinking skills are accepted as the ultimate target of educational curricula, and original training programs are formed and measurement tools are developed in this regards (Akbıyık & Seferoğlu, 2006; Butler, 2012; Karalı, 2012; Liu et al., 2014). Critical thinking has been discussed by different disciplines in its historical development and explained in more detail. Previously, philosophy

Corresponding Author: Burcu AYKAÇ **E-mail:** burcuaykac9@gmail.com
Cite this article as: Aykaç, B., & Aslan, A. E. (2022). Developing İstanbul critical thinking scale. *HAYEF: Journal of Education*, 19(1), 9-17.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
Available online at http://hayefjournal.istanbul.edu.tr/

(Ennis, 1993; Fisher & Scriven, 1997; Paul & Elder, 2013), later psychology (Halpern, 1998, 2014; Kuhn, 1999; Perkins et al., 1993), and education (Hitchcock, 2017; Ku & Ho, 2010; McPeck, 1984) fields described critical thinking based on their context and made several definitions.

The most comprehensive definition was made by a multidisciplinary expert group in the Delphi report (Facione, 1990a). According to the Delphi report (Facione, 1990a), critical thinking is a purposeful and self-regulatory judgment that originates from interpretation, analysis, evaluation and inference, and explanation of the evidence-based, conceptual, methodological, criteria-based, or contextual thoughts that make up the basis of this judgment. According to the literature, a comprehensive definition must be chosen to measure the complex and multi-component structure of critical thinking (Erwin & Sebrell, 2003; Facione, 1990a). Therefore, the scale in this study was developed based on the comprehensive and multidisciplinary definition that was created as a result of the consensus of many authorities from different disciplines and presented in the Delphi report (Facione, 1990a) in a comprehensive and integrative way.

In the Delphi report (Facione, 1990a), critical thinking is explained based on two components as "cognitive skills" and "dispositions." In this respect, a critical thinker must have critical thinking skills as well as critical thinking dispositions. According to the Delphi report (Facione, 1990a), basic cognitive skills of critical thinking include (1) interpretation, (2) analysis, (3) evaluation, (4) inference, (5) explanation, and (6) self-regulation. Basic dispositions of critical thinking in the Delphi report are divided into two dimensions as "Approaches to Life and Living in General" and "Approaches to Specific Issues, Questions, or Problems" (Facione, 1990a). To ensure the content validity in scale development studies, after determining a comprehensive definition of critical thinking, it should be clarified which aspects of critical thinking will be evaluated (Facione, 1990a). Since the purpose of the study was to develop a scale to measure cognitive performance, the basic cognitive skills in the Delphi report (Facione, 1990a) were chosen as the theoretical framework. Due to the complex structure of critical thinking skills, it was decided that the scope of the scale should be restricted to cognitive skills rather than measuring both skills and dispositions.

Critical Thinking and Critical Thinking Measurements in Higher Education

Good thinking skills are not a spontaneous phenomenon. It is necessary to provide opportunities for teaching and developing these skills. Providing education based on remembering information in secondary and post-secondary institutions and conducting knowledge-based assessments in the same way in exams often results in students not being able to develop high-level thinking skills such as critical thinking and creative thinking. Although an approach based on presenting more information is dominant in the modern education system, it is seen that many university graduates have difficulties in their business life when faced with the fact that the learned information is rapidly losing its currency. Therefore, teaching critical thinking, which is one of the most important good thinking skills sought today, should be a priority of universities throughout higher education (Beyer, 1987).

Critical thinking is given importance worldwide (Ku, 2009), and in educational reports prepared by many countries, critical thinking has been evaluated as a fundamental area that should be developed and measured in higher education (American Association of Colleges and Universities, 2005; Australian Council for Educational Research, 2002; Higher Education Quality Council, Quality Enhancement Group, 1996). Many universities agree that developing students' critical thinking skills is one of the most basic goals (Aretz et al., 1997). As the importance given to critical thinking in higher education has increased, how to measure critical thinking in universities has also become an important issue (Erwin & Sebrel, 2003).

The attempt to measure critical thinking requires reflection on two different views about the nature of thinking and logic (O'Hare, 2005). First, thinking and logic are general cognitive processes that can be transferred to different topics and contexts. (e.g. California Critical Thinking Scale–Facione, 1990b; Watson-Glaser Critical Thinking Scale–Watson & Glaser, 1980). Second, thinking and logic must be evaluated in the context of knowledge and discipline (e.g. Mathematical Packages Test–Educational Testing Service, 1994). Many widely used critical thinking scales that are suitable for university students (e.g. California Critical Thinking Scale–Facione, 1990b; Watson-Glaser Critical Thinking Scale–Watson & Glaser, 1980) were developed according to the former. Similarly, the scale in this study is developed accordingly to the first view that accepts critical thinking as a general and transferable ability.

There are various criticisms in the literature about widely used measurement tools developed to examine critical thinking abroad. The main criticism is arguing that none of the scales designed to measure critical thinking measures all of the critical thinking dimensions and components (Dogan, 2013; Erwin, 2000; Facione, 1990a; NPEC, 2000). For this reason, if these scales are used, only certain dimensions of critical thinking will be measured (Doğan, 2013). Similarly, Ennis's (1993) most basic criticism of existing measurement tools is about their insufficient inclusiveness. On the other hand, in particular, multiple-choice critical thinking tests may be insufficient due to the differences between the background beliefs and assumptions of the test maker and the test taker (Norris & Ennis, 1989). Since the concept of critical thinking is complex and there are different definitions of it, the tools developed to measure it are also diverse. However, due to the complexity of the concept and the insufficient scope of the existing measurement tools, it may sometimes be necessary to use more than one measurement tool to make a valid measurement (Aretz et al., 1997).

Critical Thinking Measurements in Turkey

Some of the critical thinking scales used in Turkey were adapted from foreign literature (Aslan et al., 2019; Aybek & Çelik, 2007; Çıkrıkçı, 1993; Demircioğlu, 2012; Dönmez & Kaya, 2016; Evcen, 2002; İskifoğlu & Ağazade, 2013; Kökdemir, 2003; Kılıç & Şen, 2014; Kurnaz, 2007; Şenturan, 2006), and some of them are originally developed scales of critical thinking (Akbıyık, 2002; Aybek et al., 2015; Görücü, 2014; Karadağ et al., 2017; Sarıgöz, 2014; Semerci, 2000, 2016; Özelçi & Saracaloğlu; 2017; Özdemir, 2005). When adaption studies are examined, some obvious drawbacks are seen. Considering the statistical and psychometric results of the adaptation studies on critical thinking scales in Turkey, it is seen that the validity and reliability values of the adapted forms are quite insufficient. Even if the scale adaptation steps are followed meticulously, findings show low validity and reliability values because of the complex nature of critical thinking, negative effects of cultural differences, and inability to provide linguistic equivalence as effectively as the original language. When scale development studies are examined, as similar to criticism in foreign literature, it is seen that developed scales to measure critical thinking do not measure critical thinking in a way that covers all its dimensions. In addition, developed scales in Turkey are generally aimed to measure critical thinking disposition or attitudes, and as in line with the foreign literature, they are constructed in Likert-type questionnaires and are based on self-report (Doğan, 2013). Therefore, the structure of these scales is not suitable to measure performance. It is determined that there is only one performance scale developed for fifth graders (Eğmir, 2016), and there is not any performance scale to measure the level of critical thinking skills of university students in Turkey.

The Aim of the Study

As seen in the literature, adaption studies of standardized measurement tools widely used are not always efficient (Doğan, 2013; Vural & Kutlu, 2004). Developing a scale that measures all dimensions of critical thinking may enable more effective studies on critical thinking in Turkey. In addition, Atkinson (1997, p. 89) claims that "critical thinking is cultural thinking". Therefore, while planning studies on critical thinking, more productive results can be obtained by considering the cultural context. Thus, there is a significant need to develop local, valid, and reliable measurement tools that can be used to measure critical thinking skills in Turkey (Alagözlü & Saraç-Süzer, 2010; Doğan, 2013; Vural & Kutlu, 2004). It is thought that a scale that will be developed in this study will meet this need and constitute a basis for future studies. Based on these reasons, this study aims to develop a local, valid, and reliable performance scale to determine the critical thinking levels of university students.

Method

The study was designed in line with the psychological test development method. As the population, university students were targeted, and different types of scores (numerical, equal-weight, verbal, and language) that were taken as a basis for university entrance were accepted as a cluster, and samples were taken from each. The data were collected by obtaining permission slips from the Social and Humanities Ethics Committee of Istanbul University-Cerrahpaşa (No. 25.10.2019-52145) and all departments through the rectorate. The data were collected from volunteer students studying in different faculties and departments of Istanbul University (N=225) and Istanbul University-Cerrahpaşa (N=805) in the 2019–2020 academic year. The distribution of students in the sample according to departments was as follows; Computer and Teaching Technologies Education (N=35, 3.4%), Turkish Teaching Department (N=63, 6.1%), French Teaching Department (N=21, 2.0%), German Teaching Department (N=46, 4.5%), Special Education Department (N=46, 4.5%), Classroom Teaching Department (N=29, 2.8%), Mathematics Teaching Department (N=32, 3.1%), Physical Education Department (N=20, 1.9%), Preschool Department (N=32, 3.1%), Journalism Department (N=20, 1.9%), Public Relations (N=9, 0.9%), and Radio-TV-Cinema (N=8, 0.8%). The distribution of the students in the sample according to grades was as follows: 501 (48.5%) freshmen, 202 (19.6%) sophomores, 112 (10.9%) juniors, and 214 (20.7%) seniors. The distribution of the students in the sample according to gender was as follows: N_{female}=706 (68.4%) and N_{male}=323 (31.4%).

Study Process

Literature Review

To develop a unique measurement tool, this study was started by a comprehensive and detailed literature review of critical thinking and developing measurement. Critical thinking cognitive skills presented in the Delphi report (Facione, 1990a) constitute the basic framework of the study.

Focus Group Interviews

This method is used to examine the knowledge and experiences of people, not only to investigate what people think but also how and why they think in this way (Kitzinger, 1995). Within the scope of this study, focus group interviews were held with a group of university students in the target group and a group of faculty members who teach at the university. The participants' perspectives, experiences, feelings, and emotions on the subject were collected under the main theme of "what is critical thinking and how to measure it". The statements of the participants were interpreted to create the content validity of the scale.

Creating the Item Pool

After the preliminary preparations, it has been decided how the scale will be constructed and what kind of questions will be asked. Accordingly, it was decided to create separate sections for each critical thinking skill and write multiple-choice items based on the evaluation of paragraphs, statements, or arguments following the purpose of the relevant section. Although measuring critical thinking only in the context of evaluation of statements and arguments causes some components of the concept to be neglected, it is difficult to go beyond this when it comes to measuring critical thinking skills through a multiple-choice test. When widely used standardized tests (e.g., Watson-Glaser, Ennis, Facione, and Fisher) of critical thinking are examined, it is seen that the following skills are aimed to be measured through multiple-choice items (Hitchcock, 2017):

- Evaluation of the inferences for a certain result from the given statements,
- Identifying an implicit assumption in a particular statement or argument,
- Clarifying the meaning,
- Evaluation of the credibility of an argument,
- Analysis of the argumentation structure in a paragraph,
- Evaluating whether the following conclusion can be reached from the given information,
- Reasoning about how to evaluate a given argument, and
- Identification of fallacies.

Accordingly, it was decided to create multiple-choice items based on examining paragraphs and statements in the context of critical thinking. Hereby, an item pool consisting of five dimensions (interpretation, analysis, evaluation, inference, and explanation) and 160 multiple-choice items were created to measure critical thinking skills. The purposes and instructions of all sections are given at the beginning of each section. Some sections have short paragraphs and various statements about them. In some other sections, however, the participant is asked to examine a series of statements given in line with the purpose of that section. Each question has three answers option and only one of these options is the correct answer. The topics of the paragraphs and statements were chosen from the topics that could appeal to university students in Turkey. An answer key was also created for the item pool.

Target Group Interviews

It is important to determine whether the responses given by the participants reflect their critical thinking levels in real terms to ensure structural validity. One of the effective ways to achieve this is to interact with the target group directly and test the structural validity with the responses given (Facione, 2000; Norris, 1988). In this way, factors avoiding the purpose of measurement, such as whether the expressions in the items are confusing, whether they are understood correctly by participants, whether background beliefs play roles in answering the question instead of the ability to think critically can be detected (Norris, 1988).

A total of six (N_{male}=3) university students that represented different grades and score types participated in the target group interviews. Students were asked to write their opinions on the evaluation form. Notes were taken by the researcher meanwhile. Some items were excluded from the scale, some others were corrected, and a new form consisting of 76 items was created to be used in the pilot study.

Pilot Study

The pilot study is conducted to increase the validity of the scale, to try the scale before the main study, and to eliminate deficiencies if there is any (Facione, 2000). For this purpose, the item pool was applied

to the students at the Science Teaching (N=35), English Teaching (N=34), and Guidance and Psychological Counseling (N=22) departments of Hasan Ali Yucel Faculty of Education. Students were asked to answer the questions and write their opinions and suggestions on the clarity and layout of the scale. Timing for the scale was determined according to the average of the students' response times.

The Final Form of the Item Pool

After the pilot study, the feedback received from the students and the notes taken by the researcher were examined and the final form was created. The final form consisted of a total of 76 items under five dimensions as interpretation (16 items), analysis (24 items), evaluation (16 items), inference (12items), and explanation (8 items). Each part of the scale is designed to measure a different critical thinking cognitive skill. Sample items are shown in Appendix 1.

Data Collection

After the necessary permissions were obtained, the final form of the item pool was conducted on the days and hours determined by making appointments with the faculties and departments. The data were collected from the Pharmacy, Law, and Communication faculties of Istanbul University, and the Health Sciences, Hasan Ali Yücel Education faculties of Istanbul University-Cerrahpaşa.

Analysis

The data obtained from 1031 students were transferred to SPSS 21 Program. As factor analyses are sensitive to missing data, missing observations in the dataset were also removed. Analyses were made on the answers of the remaining 853 students.

According to Gorsuch (1997), if there is a hypothesis about the dimensions under which the items will be placed and the relationship of the items with these dimensions, it is sufficient to use only confirmatory factor analysis (CFA). Within the scope of this study, it was aimed to determine how the items created according to an existing theoretical structure work under the dimensions in which they are included and how they work together. In other words, it was tried to verify that the items work as designed and to ensure the highest fit by removing the items that do not fit the model. Therefore, it is decided to use CFA.

CFA was performed using the Lisrel program. Bandalos and Finney (2010) suggest using weighted least square and cross-weighted least square (DWLS) methods as estimation methods for data with less than five responses categories. Within the scope of the study, DWSL (diagonally weighted least squares) was used as the estimation method. The standardized path coefficients and fit indices obtained by the analyzes were examined, and some items were gradually removed from the scale. The reliability of the total and sub-dimensions of the scale was calculated by the stratified Cronbach alpha method. The very final form of the scale was created by presenting all the results related to the structure obtained. Mean- and variance-adjusted weighted least squares (WLSMV) were recommended to predict the CFA model parameters with categorical variables (Muthén & Muthén, 2010). Multiple evaluations were made in line according to CFI > 0.90 (Browne & Cudeck, 1993), TLI > 0.90 (Bentler & Bonett, 1980), and RMSEA < 0.08 (Jöreskog & Sörbom, 1993) criteria regarding the model-data agreement.

Results

Confirmatory factor analysis results based on the diagonally weighted least squares (DWLS) estimation method were used in the analysis. As a result of many CFAs, modification indices were examined, and items that had standardized path coefficients below 0.30 were removed from the scale gradually, and the resulting three-dimensional scale of 25 items was developed (Figure 1). The developed scale was named as "İstanbul Critical Thinking Scale."

It is not enough to decide on the structure of the scale based on only statistical analysis alone. The statistical results and the items of the scale should be considered together (Erkuş, 2007). Based on this, not only the values obtained based on the analysis results were considered in the elimination of the items but also the cases such as most of the items in the item pool depend on a specific question root, that is, there are more than one question linked to a paragraph that were also taken into account. For example, the standardized path coefficient values of some items (interpretation: item 6, item 14; inference: item, 57) are just above the acceptable value (0.30) and are at a limit that can be considered low compared to some other items. But these items, which were thought to be appropriate to be included together due to the question root that they belong to, were kept in the scale.

"m" means item, "yorumlam" is for the dimension of interpretation, "degerlen" is for the dimension of evaluation, "cikarim" is for the dimension of inference.

As a result of the CFA analysis of the Istanbul Critical Thinking Scale, it was found that the standardized path coefficients of the items with three-dimensional structure as interpretation (7 items), evaluation (11 items), and inference (7 items) ranged between 0.310 and 0.641 and were statistically significant (Figure 1; Table 1).

The correlations between sub-dimensions were also examined. It was found that the highest correlation was between interpretation and evaluation (0.574, $p < .05$), and the lowest correlation was between interpretation and inference (0.385, $p < .05$). A correlation was also found between evaluation and inference (0.448, $p < .05$). All correlation values were significant at $p < .05$. The moderate correlations between dimensions were considered to be an indication of the collectability of the dimensions. When the *goodness of fit indices* given in Table 2 were examined, it was seen that the data have a good level of fit to the tested structure.

Cronbach alpha reliability values of the sub-dimensions of evaluation and inference of the scale were 0.505, 0.581, and 0.464, respectively. The overall reliability of the scale was calculated to be 0.676 with the stratified alpha. Cronbach et al., (1965) recommended using the stratified Cronbach alpha coefficient for the reliability of combined scores of sub-dimension measuring instruments. The stratified Cronbach alpha coefficient was calculated in the R program using the "sirt" package (Robitzsch, 2017).

Discussion, Conclusion, and Recommendations

This study aims to contribute to the studies on critical thinking by developing a local and original measurement tool to measure the critical thinking levels of university students, and as a result, a scale consisting of three dimensions (interpretation, evaluation, and inference) and a total of 25 items were developed. A series of steps that were described in detail under the title of "Study Process" was followed in the development of the Istanbul Critical Thinking Scale. The most important thing to be considered for content validity is to clarify which cognitive skills will be included in the scope of the scale while designing the measurement tool (Facione, 1990a; Norris, 1988). To ensure the content validity of the scale, in the first stage, a literature review was made and the scope was determined for the concept of critical thinking by researching what critical thinking is and how it is measured. In the second stage, focus group studies, which were stated to be an effective method in ensuring content validity (Kim et al., 1999; Lee & Lim, 2008; O'Brien, 2013; Powell et al., 1996; Vogt et al., 2004), with two separate groups consisting of instructors and students were conducted. Finally, after the item pool was created, five academicians who are experts in the field evaluated the relevant items in terms of content validity.

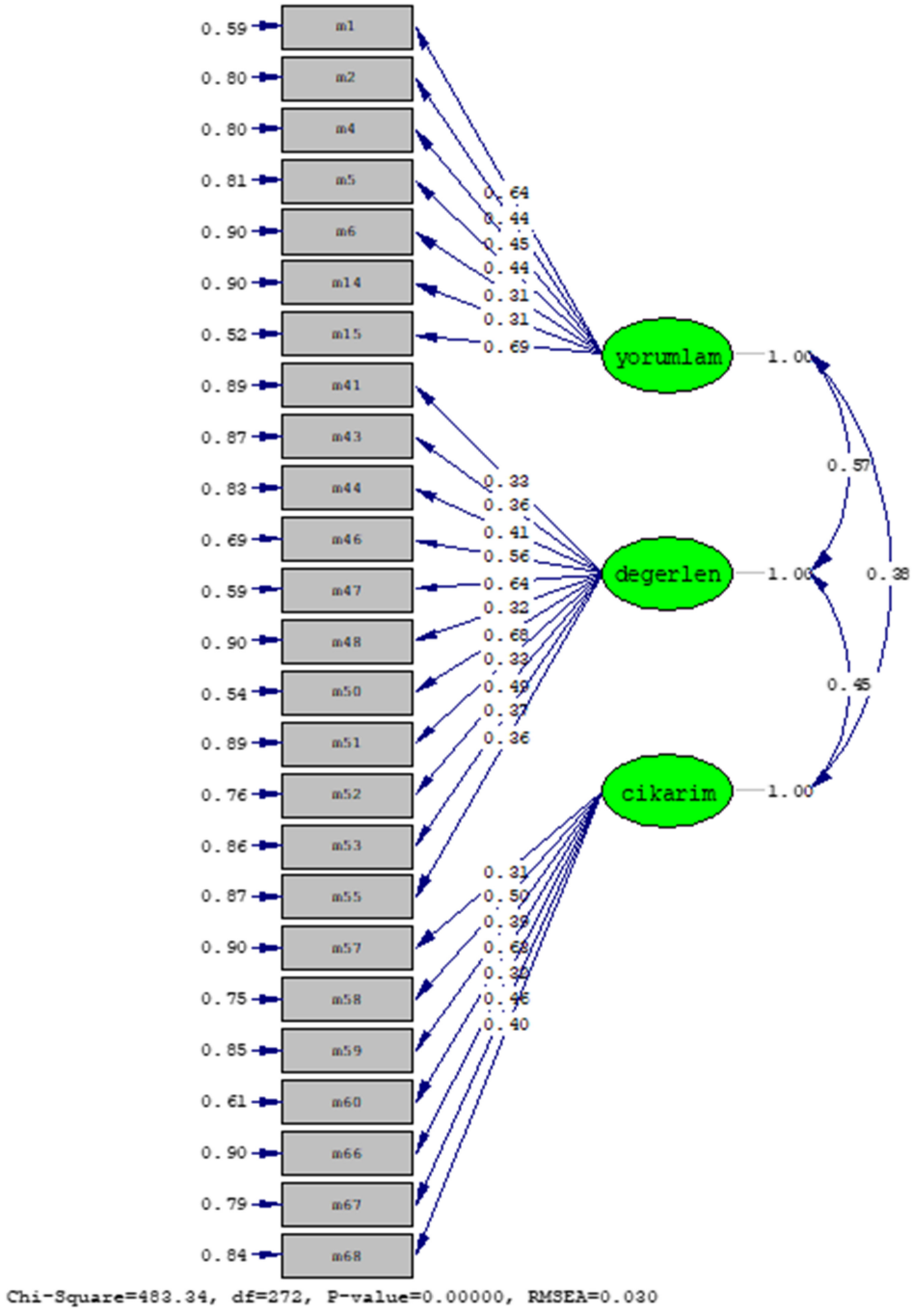

Figure 1.
Istanbul Critical Thinking Scale—ISCTS Path Diagram.

To ensure the construct validity of the scale, it is recommended to examine the item pool with a small group that has the characteristics of the target group through discussions/interviews (Facione, 1989, 2000; Norris, 1988). A group of university students was interviewed, and the necessary changes were made in the item pool by examining factors such as whether the items were understood accurately and whether background beliefs played roles in answering the questions instead of critical thinking skills. In addition, a pilot study is also recommended before the application to increase its validity (Cai, 2013; Cohen & Swerdlik, 2010; Erkuş, 2016; Lee & Lim, 2008). It is possible to identify potential problems in items and receive feedback on unclear and incomprehensible points about the item pool (Lee & Lim, 2008). The aim of the pilot study was to create the most ideal form of the scale by taking the opinions and suggestions of the participants regarding fluency, comprehensibility, and page layout.

Norris (1988) emphasizes that a researcher who develops a multiple-choice scale to measure critical thinking should use technical terms cautiously, as well as ensure the scope and structure validity. Considering that the technical words used may have an impact on test performance, it is necessary to be sensitive that the skill is not measured in a vocabulary-dependent manner (Facione, 1990a; Norris, 1988). In this respect, descriptive instructions were created for each section to ensure that knowing or not knowing the concepts specific to critical thinking does not affect the measurement results. During the target group interview and the pilot application, the clarity and the comprehensibility of the instructions/items, especially considering the technical expressions related to critical thinking, were examined. As a result of these evaluations, necessary arrangements and/or additions were made to prevent the negative effects of technical terms specific to critical thinking.

Table 1.
Standardized Path Coefficients

	Dimensions		
Items	**Interpretation**	**Evaluation**	**Inference**
Item 1	0.638	–	–
Item 2	0.443	–	–
Item 4	0.447	–	–
Item 5	0.436	–	–
Item 6	0.310	–	–
Item 14	0.311	–	–
Item 15	0.689	–	–
Item 41	–	0.333	–
Item 43	–	0.356	–
Item 44	–	0.413	–
Item 46	–	0.557	–
Item 47	–	0.641	–
Item 48	–	0.322	–
Item 50	–	0.677	–
Item 51	–	0.328	–
Item 52	–	0.488	–
Item 53	–	0.372	–
Item 55	–	0.362	–
Item 57	–	–	0.315
Item 58	–	–	0.495
Item 59	–	–	0.392
Item 60	–	–	0.627
Item 66	–	–	0.321
Item 67	–	–	0.459
Item 68	–	–	0.396

The fourth problem encountered while measuring critical thinking is to answer questions correctly or incorrectly based on having background knowledge or not. To avoid this problem, it is recommended that items should be created about social and human experiences that can be considered universal as much as possible (Facione, 1989). For this reason, while creating the item pool, attention was paid to both creating items in such a way that having/not having background information would not be an advantage/disadvantage and that the topics were chosen from daily life as much as possible. In addition, considering that it is a performance test that requires attention and will take a long time, the paragraphs and the statements were tried to be chosen from the topics that will attract the attention of the target group and they can read without getting bored.

The first designed item pool contained 76 items under five dimensions as interpretation, analysis, evaluation, inference, and explanation. CFA, RMSEA, GFI, NNFI, CFI SRMR, and x^2/df values from fit indices obtained concerning the scale were examined. As a result of the analysis, since all the standardized path coefficient values of the items in the two sub-dimensions (analysis and explanation) are much lower than 0.30, which is considered to be the minimum level, and the items negatively affect the goodness of fit indices, all items in these two sub-dimensions were removed from the scale. It is seen that the Istanbul Critical Thinking Scale with three sub-dimensions (interpretation, evaluation, and inference) consisting of the remaining 25 items shows a good fit index. The fact that these values are in the expected ranges indicate that the construct validity of the scale is provided (Byrne, 2016; Çokluk et al., 2012; Hu & Bentler, 1999; Shevlin & Miles, 1998).

Table 2.
The Goodness of Fit Indices

The Goodness of Fit Index	**Acceptable Limit***	**Value**
X^2/SD	<5 moderate <3 good	483.336/2725=1.776
CFI	>0.90	0.966
NNFI	>0.90	0.962
GFI	>0.90	0.966
RMSEA	<0.08	0.030

*Source: Çokluk, et al. (2012, pp. 271-272).
Note: X^2/SD, model chi-square; CFI=comparative fit index; NNNFI=non-normed fit index; GFI=goodness of fit; RMSEA=root mean square error of approximation.

Although all technical measures were taken, as a result of the CFA, acceptable statistical support was not found in the analysis and explanation dimensions. The reasons for this may be listed as the poor critical thinking skills of university students, the inability to develop skills such as thinking and critical thinking in the education system at a targeted level, the lack of courses to develop critical thinking (e.g. philosophy) in the educational levels up to the university, the inadequacy of teaching techniques and methods to support critical thinking skills, and living in a social and cultural environment that does not support the development of critical thinking skills. As Facione (2000) stated, the participants may not want to solve some sub-dimensions, or they may be more oriented toward others. Therefore, the psychometric values of the two sub-dimensions were low, causing them removed from the scale.

According to Ennis (2018), multiple-choice tests may not be comprehensive enough to measure critical thinking, and the differences in beliefs of test makers and test-takers may have a negative effect on the students' scores. In this study, such differences in beliefs may have caused inconsistency in some questions that students have answered. However, feedbacks that were received by the target group interviews and pilot study suggest that they had difficulty in answering questions because they were not familiar adequately with such questions and concepts that were specific to critical thinking (e.g., premises and arguments). For this reason, the questions in eliminated sections may not have been answered, or perhaps students may have marked the choices randomly without thinking. Alagözlü and Saraç-Süzer (2010) revealed in their study that being unfamiliar with the subject of argument creates a disadvantage in terms of determining critical thinking skill levels. They pointed out that the issue of argument should be embedded in educational programs and valid tools to measure critical thinking skills need to be developed.

Some skills in different dimensions of critical thinking are intertwined with each other. This situation makes it difficult to measure critical thinking dimensions by separating them from each other and may negatively affect the reliability and validity values of the scale (Dogan, 2013). The dimensions of the developed scale also contain items related to skills intertwined with each other. There is a possibility that the items were not in the correct dimension. Therefore, the removed items can be examined and tried again in prospective studies.

Cronbach alpha is usually used when the internal coefficient of consistency is calculated for the total and sub-dimensions of the scale (Güngör, 2016; Karakoç & Dönmez, 2014). In this study, the Cronbach alpha coefficient value was used in calculating the internal consistency as the answers were scored in a binary form as 1-0. The generally accepted Cronbach alpha reliability coefficient is .70 (Güngör, 2016; Karakoç & Dönmez, 2014; Lee & Lim, 2008; Nunally, 1978).

When the Cronbach alpha internal consistency coefficients are considered for the sub-dimension and sum of the developed scale, it was found that it was below the generally accepted level (0.70). This can be considered as a violation of its validity. However, due to the sensitivity of complex skill measurements such as critical thinking, it is emphasized that all

dimensions of the scale can be accepted as Cronbach alpha value of 0.67 (Norris, 1989). On the other hand, it is emphasized that the Cronbach alpha value might be lower because the sub-dimensions of the scale contain fewer items than the sum of the scale (Cronbach, 1990; Nunnally, 1978). Facione (2000) also emphasized that participants who took a test aiming to measure several different cognitive sub-skills might not do well in some parts while they could do well in others and this might decrease reliability. Based on all these facts, it is considered that the structure can be acceptable since the total reliability of the developed scale is 0.68.

Finally, the fact that the sub-dimensions of the scale were found to be positively related to each other at moderate levels and the standardized path coefficient values were above 0.30, Büyüköztürk, 2011 indicated that the Istanbul Critical Thinking Scale is acceptable.

As a result of this study, the "Istanbul Critical Thinking Scale" consisted of a total of 25 items under three dimensions as interpretation (7 items), evaluation (11), and inference (7) was developed to measure the critical thinking skills of university students. Analyses showed that the total score of the scale could be calculated since there was a moderate relationship between the dimensions. It was concluded that the scale was valid because the values obtained as a result of the CFAs showed compliance with the relevant fit indices.

As a result of analyses on reliability, it was found that the reliability of the scale was at an acceptable level. According to these results, the developed scale can be used to determine the critical thinking levels of university students including interpretation, evaluation, and inference skills. The highest score that can be obtained from the scale is 25, and the cut-off point regarding which score ranges represent low, medium, or high critical thinking skills will be determined later in the framework of additional studies on different samples. Since the critical thinking scales in the domestic literature are based on self-report and for measuring critical thinking dispositions and attitudes, this performance scale for measuring critical thinking skills is thought to make a unique contribution to the literature.

To carry out the scale development study to measure critical thinking more effectively, a program to teach critical thinking skills based on identifying and evaluating arguments to the target group can be prepared, and by applying the developed item pool as pre-test and post-test, new findings regarding the validity of the scale can be obtained. Non-functioning sub-dimensions of the five dimensions item pool can be revised and retested in a group previously defined as successful—unsuccessful by another critical thinking test.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – B.A.; Resources – A.E.A.; Materials – B.A.; Data Collection and/or Processing – B.A.; Analysis and/or Interpretation – B.A.; Literature Search – B.A.; Writing Manuscript – A.E.A.; Critical Review – A.E.A.

Conflict of Interest: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

References

Akbıyık, C. (2002). *Eleştirel düşünme eğilimleri ve akademik başarı* [*Critical thinking dispositions and academic achievement*] [Yayımlanmamış yüksek lisans tezi]. Ankara, Turkey: Sosyal Bilimler Enstitüsü, Hacettepe Üniversitesi. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi-adresinden-edinilmiştir.

Akbıyık, C., & Seferoğlu, S. S. (2006). Eleştirel düşünme eğilimleri ve akademik başarı [Critical thinking dispositions and academic achievement]. *Çukurova University Faculty of Education Journal*, *3*(32), 90–99.

Alagözlü, N., & Süzer, S. S. (2010). Language and cognition: Is critical thinking a myth in Turkish educational system? *Procedia - Social and Behavioral Sciences*, *2*(2), 782–786. [CrossRef]

American Association of Colleges and Universities. (2005). *Liberal education outcomes: A preliminary report on student achievement in college*. Washington, DC, USA: AAC&U.

Aretz, A. J., Bolen, M. T., & Devereux, K. E. (1997). Critical thinking assessment of college students. *Journal of College Reading and Learning*, *28*(1), 12–23. [CrossRef]

Aslan, A. E., Bozdağ, F., & Aykaç, B. (2019). Turkish validity and reliability of Watson-Glaser Critical Thinking Appraisal Test's Short Form. In *Proceedings of the International Psychological Applications Conference and Trends*. [CrossRef]

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly*, *31*(1), 71–94. [CrossRef]

Australian Council for Educational Research (2002). *Graduate skills assessment*. Australia: Commonwealth of Australia.

Aybek, B., Aslan, S., Dinçer, S., & Arısoy, B. C. (2015). Öğretmen adaylarına yönelik eleştirel düşünme standartları ölçeği: Geçerlik ve güvenirlik çalışması [Critical thinking standards scale for prospective teachers: Validity and reliability study]. *Kuram ve Uygulamada Eğitim Yönetimi Dergisi*, *21*(1), 25–50.

Aybek, M., & Çelik, M. (2007). Watson Glaser eleştirel akıl yürütme gücü ölçeği'nin (w-geaygö) üniversite ikinci üçüncü ve dördüncü sınıf İngilizce bölümü öğretmen adayları üzerindeki güvenirlik çalışması [Reliability study of Watson Glaser's critical appraisals (w-gcta) on second, third and fourth year university English department teacher candidates]. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *16*(1), 101–112.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 98–122). New York, NY, USA: Routledge.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. [CrossRef]

Beyer, B. K. (1987). *Practical strategies for the teaching of thinking*. Boston, MA, USA: Allyn and Bacon.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA, USA: SAGE Publications.

Butler, H. A. (2012). Halpern critical thinking assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, *26*(5), 721–729. [CrossRef]

Büyüköztürk, Ş. (2011). *Sosyal bilimler için veri analizi el kitabı* [*Manual of data analysis for social sciences*]. Ankara, Turkey: Pegem Akademi.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed). New York, NY, USA: Routledge.

Cai, L. (2013). Factor analysis of tests and items. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology volume 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 85–100). Washington, DC, USA: American Psychological Association. [CrossRef]

Cohen, R. J., & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement*. New York, NY, USA: McGraw-Hill.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed). New York, NY, USA: Harper & Row.

Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement*, *25*(2), 291–312. [CrossRef]

Çıkrıkçı, N. (1993). Watson-Glaser eleştirel akıl yürütme gücü ölçeğinin (form ym) lise öğrencileri üzerindeki ön deneme uygulaması [Pretesting of Watson-Glaser critical thinking appraisal scale (form ym) on high school students]. *Ankara University of Journa of Educational Sciences*, *25*(2), 559–569.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (cilt. *2*) [*Multivariate statistics for social sciences: SPSS and LISREL applications* (2nd ed)]. Ankara, Turkey: Pegem Akademi.

deLantante, K. (2019). *The critical thinker academy, Critical thinking: What it is and why it's important* [Online Course]. Retrieved from https://criticalthinkeracademy.com/p/what-is-critical-thinking.

Demircioğlu, E. (2012). *Eleştirel düşünme eğilimi ölçeğinin uyarlama çalışması ve faktör yapısının farklı değişkenlere göre incelenmesi* [*Adaptation study of the critical thinking disposition scale and examining the factor structure according to different variables*] [Yayımlanmamış yüksek lisans tezi]. Bolu, Turkey: Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresindenedinilmiştir.

Diker-Coşkun, Y. (2018). Yenilik çağında düşünme: Neden eleştirel düşünmeliyiz? [Thinking in the age of innovation: Why should we think critically?]. In D. B. Yenici (Ed.), *Eleştirel düşün: Öğren, sorgula, üretiçinde* [*Think critically: learn, question, produce*] (ss. 13–15). İstanbul, Turkey: Yeditepe Üniversitesi Yayınevi.

Doğan, N. (2013). Eleştirel düşünmenin ölçülmesi [Measuring critical thinking]. *Cito Education: Theory into Practice*, *22*, 29–42.

Dönmez, B., & Kaya, F. (2016). Eleştirel düşünme motivasyonu ölçeği'nin Türkçe'ye uyarlanması [Turkish Adaptation study of critical thinking matıvational scale]. *HAYEF Journal of Education*, *13*(2), 159–173. Retrieved from https://dergipark.org.tr/tr/pub/iuhayefd/issue/24491/259590.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52. [CrossRef]

Educational Testing Service (1994). *Packets*. Princeton, NJ, USA: D.C.Heath and Company.

Eğmir, E. (2016). *Eleştirel düşünme becerisi öğretim programının hazırlanması, uygulanması ve değerlendirilmesi* [*Preperation, implementation and evaluation of critical thinking skill curriculum*] [Yayınlanmamış doktora tezi]. Afyonkarahisar, Turkey: Afyon Kocatepe Üniversitesi, Sosyal Bilimler Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresi ndenedinilmiştir.

Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, *32*(3), 179–186. [CrossRef]

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi*, *37*(1), 165–184. [CrossRef]

Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar [Problems encountered in scale development and adaptation studies]. *Türk Psikoloji Bülteni*, *13*(40), 17–25.

Erkuş, A. (2016). *Psikolojide ölçme ve ölçek geliştirme- I temel kavramlar ve işlemler* (baskı. *2*) [*Measurement and scale development in psychology- I basic concepts and operations* (2nd ed)]. Ankara, Turkey: Pegem Akademi.

Erwin, T. D. (2000). *The NPEC sourcebook on assessment, volume 1: Definitions and assessment methods for critical thinking, problem solving, and writing* [Unpublished manuscript]. Washington, DC, USA: AAHE.

Erwin, T. D., & Sebrell, K. W. (2003). Assessment of critical thinking: ETS's tasks in critical thinking. *Journal of General Education*, *52*(1), 50–70. [CrossRef]

Evcen, D. (2002). *Watson-Glaser eleştirel akıl yürütme gücü testinin (form s) Türkçe'ye uyarlama çalışması* [*Turkish adaptation study of Watson-Glaser critical thinking appraisal test (form s)*] [Yayınlanmamış yüksek lisans tezi]. Ankara, Turkey: Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresindenedinilmiştir.

Facione, P. A. (1989). Assessing inference skills. *Educational Resources Information Center*, Doc. no. TM 012917, 1–30.

Facione, P. A. (1990a). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations (ERIC Report No. ED315423). Retrieved from https://eric.ed.gov/?id=ED315423.

Facione, P. A. (1990b). *The California critical thinking skills test-college level technical report 1: Experimental validation and content validity*. Millbrae, USA: California Academic Press.

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

Fisher, A., & Scriven, M. (1997). *Critical thinking: Its definition and assessment*. Norwich, England: Edgepress.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68*(3), 532–560. [CrossRef]

Görücü, E. (2014). *Altıncı, yedinci ve sekizinci sınıf öğrencilerinin okuma alışkanlıkları ile eleştirel düşünme becerileri arasındaki ilişkinin incelenmesi* [*Examining the relationship between sixth, seventh and eighth grade students' reading habits and critical thinking skills*] [Yayınlanmamış yüksek lisans tezi]. İstanbul, Turkey: Yeditepe Üniversitesi, Sosyal Bilimler Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresi ndenedinilmiştir.

Güngör, D. (2016). Psikolojide ölçme araçlarının geliştirilmesi ve uyarlanması kılavuzu [A guide to the development and adaptation of measurement tools in Psychology]. *Türk Psikoloji Yazıları*, *19*(38), 104–112.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449–455. [CrossRef]

Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed). New York, NY, USA: Psychology Press.

Higher Education Quality Council, Quality Enhancement Group (1996). *What are graduates? Clarifying the attributes of "graduateness*. London, England: HEQC.

Hitchcock, D. (2017). Critical thinking as an educational ideal. In. *On reasoning and argument* (pp. 477–497). Cham, Switzerland: Springer. [CrossRef]

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

İskifoğlu, G., & Ağazade, A. S. (2013). Translation and validation of a Turkish version of the California Critical Thinking Disposition Inventory. *Social Behavior and Personality*, *41*(2), 187–196. [CrossRef]

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL, USA: Scientific Publishing Software International.

Karadağ, F., Demirtaş, Y. V., & Yıldız, T. (2017). 5–6 yaş çocuklar için felsefi sorgulama yoluyla eleştirel düşünmenin değerlendirilmesi ölçeğinin geliştirilmesi [Developing a scale for evaluating critical thinking through philosophical inquiry for children aged 5–6]. *International Online Journal of Educational Sciences*, *9*(4), 1025–1037.

Karakoç, A. G. D. F. Y., & Dönmez, P. D. L. (2014). Ölçek geliştirme çalışmalarında temel-ilkeler [Basic principles in scale development studies]. *Tıp Eğitimi Dünyası*, *13*(40), 39–49. [CrossRef]

Karalı, Y. (2012). *Eğitim fakültesi öğrencilerinin eleştirel düşünme eğilimleri: İnönü Üniversitesi örneği* [*Critical thinking dispositions of education faculty students: Sample of İnönü University*] [Yayınlanmamış yüksek lisans tezi]. Malatya, Turkey: İnönü Üniversitesi, Eğitim Bilimleri Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresindenedinilmiştir.

Kılıç, H. E., & Şen, A. İ. (2014). UF/EMI eleştirel düşünme eğilimi ölçeğini Türkçeye uyarlama çalışması [Turkish adaptation study of UF/EMI critical thinking disposition scale]. *Eğitim ve Bilim*, *39*(176), 1–12.

Kim, B. S. K., Atkinson, D. R., & Yang, P. H. (1999). The Asian values scale: Development, factor analysis, validation, and reliability. *Journal of Counseling Psychology*, *46*(3), 342–352. [CrossRef]

Kitzinger, J. (1995). Qualitative research: Introducing focus groups. *BMJ*, *311*(7000), 299–302. [CrossRef]

Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA, USA: SAGE Publications.

Kökdemir, D. Y. (2003). *Belirsizlik durumlarında karar verme ve problem çözme* [*Decision making and problem solving in the cases of uncertainty*] [Yayımlanmamış doktora tezi]. Ankara, Turkey: Ankara Üniversitesi, Sosyal Bilimler Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresindenedinilmiştir.

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, *4*(1), 70–76. [CrossRef]

Ku, K. Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning*, *5*(3), 251–267. [CrossRef]

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, *28*(2), 16–46. [CrossRef]

Kurnaz, A. (2007). *İlköğretim beşinci sınıf sosyal bilgiler dersinde beceri ve içerik temelli eleştirel düşünme öğretiminin öğrencilerin eleştirel düşünme becerileri, erişi ve tutumlarına etkisi* [*The effect of skill and content-based critical thinking instruction on students' critical thinking skills, achievement and attitudes in primary school fifth grade social studies lesson*] [Yayımlanmamış doktora tezi]. Konya, Turkey: Selçuk Üniversitesi, Sosyal Bilimler Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresindenedinilmiştir.

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports*, *6*, 40–41.

Lee, D., & Lim, H. (2008). Ölçek geliştirme (Çev. D. M. Siyez). In P. P. Heppner, B. E. Wampold & D. M. Kivlighan (Eds.), *Psikolojik danışmada araştırma yöntemleri içinde* (ss. 533–570). Ankara, Turkey: Mentis Yayıncılık.

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series*, *2014*(1), 1–23. [CrossRef]

McPeck, J. (1984). The evaluation of critical thinking programs: Dangers and dogmas. *Informal Logic*, *6*(2), 9–13. [CrossRef]

Muthén, L. K., & Muthén, B. O. (2010). *Mplus: Statistical analysis with latent variables user's guide 6.0*. Los Angeles, CA, USA: Muthén & Muthén.

National Postsecondary Education Cooperative (NPEC). (2000). *Definitions and assessment methods for critical thinking, problem solving, and writing*. Washington, DC, USA: United States Department of Education.

Norris, S. P. (1988). Controlling for background beliefs when developing multiple-choice critical thinking tests. *Educational Measurement: Issues and Practice*, *7*(3), 5–11. [CrossRef]

Norris, S. P., & Ennis, R. (1989). *Evaluating critical thinking*. Pacific Grove, CA, USA: Midwest Publications.

Nunnally, J. C. (1978). An overview of psychological measurement. In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders* (pp. 97–146). Boston, MA, USA: Springer. [CrossRef]

O'Brien, K. (1993). *Improving survey questionnaires through focus groups-Successful focus groups: Advancing the etate of the art* (pp. 105–117). Thousand Oaks, CA, USA: SAGE Publications. [CrossRef]

O'Hare, L. C. (2005). *Measuring critical thinking skills and dispositions in undergraduate students* [Unpublished doctoral dissertation]. Belfast, Northern Ireland: Queen's University, Belfast Faculty of Science and Agriculture. Retrieved from https://pure.qub.ac.uk/portal/en/publications/measuring-critical-thinking-skills-and-dispositions-in-undergraduate-students(65baee6c-8a10-4adb-ba62-9dc1a1b52055).html.

Özdemir, S. M. (2005). Üniversite öğrencilerinin eleştirel düşünme becerilerinin çeşitli değişkenler açısından değerlendirilmesi [Evaluation of university students' critical thinking skills in terms of various variables]. *Journal of Turkish Educational Sciences*, *3*(3), 297–316.

Özelçi, S. Y., & Saracaloglu, A. S. (2017). The development of critical thinking attitudes scale. *European Journal of Education Studies*, *3*(6), 691–702.

Paul, R., & Elder, L. (2013). *Kritik düşünce: Yaşamınızın ve öğrenmenizin sorumluluğunu üstlenmek için araçlar* (Çev. A. E. Aslan & G. Sart). Ankara, Turkey: Nobel Akademik Yayıncılık.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly*, *39*(1), 1–21.

Powell, R. A., Single, H. M., & Lloyd, K. R. (1996). Focus groups in mental health research: Enhancing the validity of user and provider questionnaires. *International Journal of Social Psychiatry*, *42*(3), 193–206. [CrossRef]

Robitzsch, A. (2017). Package 'sirt'. Retrieved from http://r.meteo.uni.wroc.pl/web/packages/sirt/sirt.pdf.

Sarıgöz, O. (2014). Öğretmen adaylarının eleştirel düşünme becerileri hakkındaki görüşlerinin değerlendirilmesi [Opinions regarding the evulation of critical thinking skills teacher]. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, *41*, 1–15. Retrieved from https://dergipark.org.tr/tr/pub/abuhsbd/issue/32979/366626.

Schneider, M., & Liu, J. (2019). *FutureLearn, Critical thinking at university: An introduction* [Online Course]. Retrieved from https://www.futurelearn.com/courses/critical-thinking-at-university.

Semerci, N. (2000). Kritik düşünme ölçeği [Critical thinking scale]. *Eğitim ve Bilim*, *25*(116), 23–26.

Semerci, N. (2016). Eleştirel düşünme eğilimi (ede) ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik revize çalışması [Development of critical thinking disposition (ctd) scale: Validity and reliability revision study]. *Electronic Turkish Studies*, *11*(9), 725–740.

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, *25*(1), 85–90. [CrossRef]

Şenturan, L. (2006). *Hemşerilik öğrencilerinde eleştirel düşünme* [Yayımlanmamış doktora tezi]. İstanbul, Turkey: Marmara Üniversitesi, Sağlık Bilimleri Enstitüsü. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkeziadresindenedinilmiştir.

Vogt, D. S., King, D. W., & King, L. A. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment*, *16*(3), 231–243. [CrossRef]

Vural, R. A., & Kutlu, O. (2004). Eleştirel düşünme: Ölçme araçlarının incelenmesi ve bir güvenirlik çalışması [Critical thinking: Examination of measurement tools and a reliability study]. *Journal of Çukurova University Institute of Social Sciences*, *13*(2), 189–200.

Watson,G., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal*. San Antonio, TX, USA: Psychological Corp.

Appendix 1

Sample Items of Istanbul Critical Thinking Scale

This test, which was developed to measure your critical thinking skills, consists of three parts in total. Each section is designed to measure a different critical thinking cognitive skill. At the beginning of each section, there is the purpose and instruction of that section. The test contains a total of 25 multiple-choice questions. These questions consist of various paragraphs and/or statements. Each question has three options and only one of these options is the correct answer. When answering the questions, please read the explanations provided to you carefully. You can show the answer you think is correct by painting the circle inside the relevant option () or marking it with a cross (X). If there is an option you want to change, please delete the previous one and/or highlight the new option. You can answer the questions in any order you want. Please do not leave blank questions/sections in terms of the efficiency of research results.

Thank you for your participation.

(1) INTERPRETATION

Purpose:
The aim of this section is to measure the interpretation skill, which is one of the critical thinking cognitive skills. Interpretation is to grasp and express the meaning or significance of a wide variety of experiences, situations, knowledge, beliefs, and judgments.

Instruction:
There are some paragraphs in this section. Some statements regarding the paragraph are presented under each paragraph. Here you are asked to read the paragraphs and statements carefully and to comment on whether the statements can be reached within the framework of the information presented in the paragraphs.

• If you think this statement will be reached based on what is explained in the paragraph, choose the **"Conclusion"** option.
• If you think that there is no information regarding this statement in the paragraph, + the **"Insufficient Information"** option.
• If you think that this statement cannot be reached based on what is explained in the paragraph, check the **"Not Conclusion"** option.

Paragraph 3:

Many consumers are concerned about the possible side effects of artificial sweeteners, as it is highly controversial whether sweeteners are healthy to use. While some experts support the use of artificial sweeteners, others argue that they are not as innocent as they seem. For example, some experts think that artificial sweeteners may be a good way to reduce sugar intake because they do not contain calories. Some experts believe that the use of artificial sweeteners may cause weight gain, not weight loss, even if it does not contain calories. So the answer to the question of whether artificial sweeteners are safe will likely depend on who you ask.	**Conclusion**	**Insufficient Information**	**Not Conclusion**
6. The fact that it does not contain calories does not guarantee that artificial sweeteners are healthy.	o	o	o
7. It is agreed by experts that artificial sweeteners do not help healthy weight loss.	o	o	o

(2) EVALUATION

Purpose:
The aim of this section is to measure assessment skills, which is one of the critical thinking cognitive skills. Evaluation is the evaluation of the logical strength of statements presented to explain a belief, opinion, judgment, or knowledge. With a view to determining the acceptability of an argument, it is to judge the evidentiary/supporting power of its premises in reaching conclusions. This section consists of two parts and the instructions for each are explained separately.

SECTION 1

Instruction:
It is necessary to evaluate whether the written or oral statements, that is, the arguments, that are presented to allow an idea, claim, or information to be accepted are valid or invalid. If the premises of an argument (reasons given to support the claim) offer logical support that guarantees to reach that claim, then that argument is valid. You can see whether an argument is valid or not without knowing the truth value (being true/false) of its premises because the validity or invalidity of the argument depends not on the truth value of those premises but on the logical relationship between the premises of that argument and its conclusion. In summary, if it is possible to conclude from the premises, that argument is valid, if not possible, that argument is considered invalid.
There are some expressions in this section. Here you are asked to read these statements carefully and first decide whether these statements are an argument. If you think statements are arguments, secondly examine their premises and conclusion. Consider whether it is possible to reach that conclusion from the premises presented, ignoring the actual degrees of accuracy of the premise and conclusion.

If you think it is possible to reach that conclusion from the presented premises, tick the **"Valid Argument"** option.
If you think it is not possible to reach that conclusion from the presented premises, please tick the **"Invalid Argument"** option.
If you think the presented statements are not arguments, check the **"Not an Argument"** option.

Please answer the questions (8–13) in this section according to the instructions above.		Valid Argument	Invalid Argument	Not an Argument
8.	Berk has two sisters. The age difference with his sisters is small. Siblings with less age difference get along better. That is why Berk gets on very well with his sisters.	o	o	o
10.	The physical images of all fraternal twins are different. Beren and Birce are fraternal twins. Therefore, the physical appearances of Beren and Birce are different from each other.	o	o	o

Instruction:
It is necessary to examine the strength of the logical link between the premises that make up an argument and the conclusion. If we assume that all the premises are true;
If it is absolutely possible to reach the conclusion that the argument tries to convince from these premises, it is called a "valid argument"; if it is uncertain but most likely, it is called a "strong argument"; if it is unlikely, it is called a "weak argument".
Various arguments are presented in this section. Here you are asked to read these arguments carefully and examine the strength of the logical link between the premises of the arguments and their consequences. Consider the extent to which it is possible to arrive at that conclusion from the premises presented, ignoring the actual degrees of accuracy of the premise and conclusion.

Check the **"Valid Argument"** option if you think it is absolutely possible to reach that conclusion from the premises presented.
If you think it is highly possible to reach that conclusion from the premises presented, tick the **"Strong Argument"** option.
If you think it is unlikely to reach that conclusion from the premises presented, tick the **"Weak Argument"** option.

Please answer the questions (14–18) in this section according to the instructions above.		Valid Argument	Strong Argument	Weak Argument
14.	There are things that many people want to go back to and change. If time travel were possible, we would have a chance to go back and change things. Since it is impossible to travel through time, none of us can change the past, even if we want to.	o	o	o
16.	Mr. Rüştü is an old person. Almost all of the elderly have a heavy hearing. People who do not hear well, volüme up the television too much. Mr. Rüştü is probably voluming up television too much.	o	o	o

(3) INTERFERENCE

Purpose:
The aim of this section is to measure inference skills, which is one of the critical thinking cognitive skills. The inference is to derive, given a set of statements, explanations, or other forms of representation, the appropriate level of logical power and their inferential relationships, the conclusions or assumptions they support, warrant, or imply.

Instruction:
This section contains various paragraphs and some inferences regarding these paragraphs. Here, you are asked to read these paragraphs and the inferences carefully and examine the truth values (being true/false) and truth degrees (definitely/possibly) of these inferences while adhering to the paragraphs. These paragraphs contain controversial issues. You are expected not to be influenced by your own opinion on the subject and to draw inferences by adhering to the paragraph.

If you think the inference presented is absolutely correct, check the **"Absolutely True**" option.
If you think the inference presented is probably correct, check the **"Probably True"** option.
Check the **"Insufficient Information"** option if you could not find any information regarding the inference presented.

Paragraph 2:

If you have entered a university in a city far from your family, you need to make a good decision about where to stay. The most advantageous of all options is to stay in a dormitory. When you stay in a dormitory, you do not have to pay additional costs such as electricity, water, natural gas, and internet. The fee you will pay for the dormitory is fixed and you will not be in a difficult situation with unexpected expenses. On the other hand, you can have a more colorful social life thanks to the dormitory. You can easily make new friends in the common areas of the dormitory and spend your student years in a community and having fun. Another good thing about staying in a dormitory is that your responsibility is less than staying at home. You will get rid of many tiring and time-consuming tasks such as cooking, washing the dishes, and cleaning the mess. This leaves you with plenty of time to study, improve, and have fun.		Absolutely True	Insufficient Information	Probably True
24.	People staying in the dormitory have a more social university life.	o	o	o
25.	Those who stay at home have to deal with tiring and time-consuming household chores.	o	o	o