KURAM VE UYGULAMADA EĞİTİM BİLİMLERİ EDUCATIONAL SCIENCES: THEORY & PRACTICE

Received: June 13, 2018
Revision received: August 15, 2018
Accepted: September 2, 2018

www.estp.com.tr
DOI 10.12738/estp.2018.3.0175 ♦ 2018 ♦ 18(3) ♦ 737–757

Research Article

Adaptation of Scientific Reasoning Scale into Turkish and Examination of its Psychometric Properties*

Gülfem Muşlu Kaygısız[1]
Hasan Kalyoncu University

Burcu Gürkan[2]
Hasan Kalyoncu University

Ufuk Akbaş[3]
Hasan Kalyoncu University

Abstract

In this study, it is aimed to adapt the Scientific Reasoning Scale (SRS) into Turkish. The translated form has been provided to the students enrolled at different levels together with a form in which they were requested to present what they have understood and the reason of their responses. It was seen that the explanations of students to one item were inconsistent or insufficient, and an alternative item was added to the test. Investigations based on PCA and CFA revealed that the psychometric properties of the test containing the alternative item were better. Moderate and positive correlations were found between the SRS scores and the CCTDI and LCTSR scores. The applications carried out before and after the course on scientific research methods carried out during a term at the undergraduate level revealed that a significant increase in the SRS scores was achieved. It was seen that there was significant difference between the scores of the students enrolled at undergraduate and graduate levels. Average item difficulty was .40 and discrimination indices was .50 and over. As result, it is seen that SRS could be used for measuring the scientific reasoning ability of the undergraduate students.

Keywords

Reasoning • Scientific reasoning • Scale adaptation • Measurement tool • Scientific Reasoning Scale

* This paper was presented at the 6th International Congress on Measurement and Evaluation in Education and Psychology, 5-8 September 2018, in Prizren, KOSOVO.

1 **Correspondence to:** Gülfem Muşlu Kaygısız, Department of Science Education, Hasan Kalyoncu University, Havaalanı yolu üzeri 8. km Şahinbey, Gaziantep Turkey. Email: gulfem.muslu@hku.edu.tr

2 Department of Curriculum and Instruction, Hasan Kalyoncu University, Havaalanı yolu üzeri 8. km Şahinbey, Gaziantep Turkey. Email: burcu.gurkan@hku.edu.tr

3 Department of Measurement and Evaluation in Education, Hasan Kalyoncu University, Havaalanı yolu üzeri 8. km Şahinbey, Gaziantep Turkey. Email: ufuk.akbas@hku.edu.tr

Citation: Muşlu Kaygısız, G., Gürkan, B., & Akbaş, U. (2018). Adaptation of Scientific Reasoning Scale into Turkish and examination of its psychometric properties. *Educational Sciences: Theory & Practice, 18,* 737–757. http://dx.doi.org/10.12738/estp.2018.3.0175

People born with their potential to think and try to figure out by classifying what they think about both themselves and the universe they live in. This is regarded as the basic function of thinking. The tendency to think is the willingness of the individual to think (Siegel, 1999); and thinking on the foundations of beliefs and their consequences is the best one (Dewey, 1910). Thinking; is shaping those obtained by observation, experience, intuition and reasoning (Özden, 2011); it is the ability to compare, distinguish, combine, connect and understand the forms (Türk Dil Kurumu [TDK], 2018). Thinking is conceptually defined in various forms; but it is very difficult to reduce the expression because of its abstract structure and depth. According to Cüceloğlu (2005), the systematic transformation of mental representations in order to describe the actual or probable state of the world is called thinking. Similarly, thinking is expressed as a systematic transformation of mental process aimed at understanding the present situation (Holyoak & Morrison, 2004). And according to Morgan (2013), thinking refers to symbolic mediation that fills the gap between the stimulating situation and the behavior it demonstrates with the inner processes individual possesses. Regarding these definitions, it is possible to explain the concept of thinking as the form of active mental tendencies people need to understand the world.

Thinking is a natural and complex activity of our mind. People think to determine the difference between facts and ideas, to criticize a thought or opinion, to create useful questions that guide a research, to solve problems, and to use results in order to predict hypotheses (such as analogical, mathematical, causal), to evaluate and see multiple perspectives, to make decisions, to assess the validity of information sources (McGuinness, 2005, Nessel & Graham, 2007). The human mind is active within the scope what it takes from external world and what it needs. Thinking process consists of logical processes called "reasoning, scientific thinking, problem solving, decision making, critical thinking, creative thinking, reading comprehension and writing" and so on (Beyer, 1988; Butterworth & Thwaites, 2013; Güneş 2012; Lipman, 2003; Özden, 2011). Reasoning is one of these thinking activities and it expresses a form of thinking based on proof; and it can also be described as treating one of the both ideas as proving the other one and reaching a conclusion (Özlem, 2014). In other words, reasoning is a human thought that supports the discovery of what is known, assumed, unknown, or implied (Barbey & Barsalou, 2009). Reasoning skills require the reasons to be based on logical consequences purified from emotions. In order for reasoning to be a subject of logic, it is necessary for the propositions to follow the order of (i) proving/premise and (ii) proved/result (Özlem, 2014). Logic is an indispensable tool of *scientific thinking*, and epistemological perspectives try to confirm facts by establishing hypotheses that reveal these facts detected by observing or experimenting. And all these actions are accepted as a logical process (Özlem, 2014; Yıldırım, 2005).

Scientific thinking is a way of thinking that has been developed by considering thought-specific constructs in any scientific context, content or problem, and by imposing intellectual standards on them. Scientific thinkers reveal vital scientific questions and problems, formulate them clearly, collect and evaluate scientific data to interpret the questions and problems that arise, test against relevant criteria and standards and finally reach to well considered scientific results and solutions (Paul & Elder, 2014). The information which is the product of the path of scientific thinking is information that may change in the context of new events and data reflecting current time. In this context, scientific thinking is a consistent and logical thinking function that the individual implements in order to solve any problem (Stuessy, 1984) and it is a cognitive process based on observations, including inductive laws, explanatory theories and hypothesis testing (Bady, 1979; Schauble, 1996). Scientific thinking also involves various general cognitive actions such as induction, deduction, analogical thinking, problem solving, and causal reasoning (Dunbar & Fugelsang, 2004). Scientific reasoning, which is described as the implementation of thinking about scientific knowledge, is also considered as a perspective of general reasoning and it is argued that it can be developed through education (Adey & Csapo, 2012; Hogan & Fisherkeller, 2005).

Kind and Osborne (2016) indicate that scientific reasoning has six types: mathematical deduction, experimental evaluation, hypothetical modeling, categorization and classification, probabilistic reasoning and history-based evolutionary reasoning. Additionally, scientific reasoning means the cognitive abilities expressed in five dimensions, namely; (i) the serial ordered reasoning, which is the ability to sort data; (ii) the theoretical reasoning required to interpret the data; (iii) functionality reasoning with the ability to analyze functional relationships; (iv) manipulating variables which is the ability to control variables and (v) probabilistic reasoning, which includes the ability to predict on the basis of data (Shofiyah, 2013 as cited in Novia & Riandi, 2017). With a similar classification, scientific reasoning deals with basic reasoning skills that enable the research process to be successful; including exploring a problem, formulating and testing hypotheses, controlling and manipulating variables, observing and evaluating experimental consequences (Han, 2013; Zimmerman, 2007). All of these explanations make it possible to say that scientific thinking, causality and general reasoning processes are the basis of scientific reasoning. In this context, scientific reasoning can be explained as disciplining the mind according to scientific research activities.

One of the most powerful features of science is to aim obtaining evidence and present that evidence as scientific knowledge. Drummond and Fischhoff (2017) argue that the ability to assess scientific evidence is necessary for most decisions to be made in individual lives. They state that scientific evidence evaluation is related to scientific reasoning skills and that scientific findings and innovations play an important role in everyday life from the technologies used to the evidence presented in political debate. Therefore, they have developed the Scientific Reasoning Scale (SRS) using an

Table 1
Information on Scales Used in the Measurement of Reasoning and Scientific Process Skills

Tool	Developed / Adapted By	Measured Trait	Application Group	Type of Item	Number of Questions Application Period
Lawson's Classroom Test of Science Reasoning	Developed by: Lawson (1978)	Conservation, proportional thinking, controlling variables, probabilistic thinking, correlational thinking, hypothetical thinking	High School (Grade 9 and 10)	Multiple choice	24 items No time mentioned
A Group Test of Logical Thinking	Developed by: Tobin and Copie (1981) Adapted by: Geban, Aşkar and Özkan (1992)	Proportional reasoning, controlling variables, probabilistic reasoning, correlational reasoning, combinatorial reasoning	High School (Grade 9)	Multiple choice Additionally written explanation is requested for the given answer.	10 items No time mentioned
Scientific Process Skill Test	Developed by: Okey, Wise and Burns (1985) Adapted by: Geban et.al. (1992)	Identifying variables, identifying and stating hypotheses, operationally defining, designing investigations, graphing and interpreting	High School (Grade 9)	No item type is mentioned	36 items No time mentioned
Group Test of Logical Thinking	Developed by: Roadrangka, Yeany, and Padilla (1982) Adapted by: Aksu, Berberoğlu, and Paykoç (1991)	Conservation, proportional thinking, controlling variables, combinational reasoning, probabilistic thinking, correlational thinking	Undergraduate	Multiple choice Open ended	21 items
Scientific Process Skills Test	Developed by: Önal (2005)	Inference, integration of information into everyday life, graphics and tables, data analysis, analysis, synthesis	Secondary School (Grade 7)	Multiple choice	36 items No time mentioned
Science Process Skills Test	Developed by: Tatar (2006)	Observation, classification, concluding, interpreting, measuring, communicating, space and time associations, stating hypothesis, experimenting, defining and controlling variables, interpreting data	Secondary School (Grade 7)	Multiple choice	18 items No time mentioned
Science Process Skills Test	Developed by: Öztürk (2008)	Observation, classification, determining variables, estimating, interpreting measurement and data, number and space relations, stating hypothesis, decision making, modeling, manipulating and controlling variables, recording data, experimenting, deduction	Secondary School (Grade 7)	Multiple choice	26 items 40 minutes
Logical Thinking Test	Developed by: Sezen and Bülbül (2011)	Defining and controlling variables, associating, calculating probability, interpreting graphics, transforming numerical expressions into graphs	Undergraduate (Mathematic education)	Multiple choice	10 items No time mentioned
Scientific Process Skill Test	Developed by: Türker (2011)	Basic scientific process skills (observing, measuring, classifying, recording data, establishing number and space relations) Causal scientific process skills (predicting, identifying variables, interpreting data, drawing conclusions) Experimental scientific process skills (stating hypothesis, using data and modeling, experimentation)	Secondary School (Grade 6)	Multiple choice	25 items 30 minutes

Table 1
Information on Scales Used in the Measurement of Reasoning and Scientific Process Skills (continued)

Tool	Developed / Adapted By	Measured Trait	Application Group	Type of Item	Number of Questions Application Period
Science Process Skills Scale	Developed by: Aydoğdu, Tatar, Yıldız, and Buldur (2012)	Basic skills (observing, classifying using space / time relations, making predictions, making inferences) High level skills (problem statement, generating hypothesis, variable determination and control, experimenting testing and data interpretation)	Secondary School (Grades 6-8)	Multiple choice	28 items No time mentioned
Science Process Skill Test	Developed by: Feyzioğlu, Demirdağ, Akyıldız, and Altun (2012)	Observing, classifying, measuring, relationship between numbers and space, predicting, identification of variables, formulating hypothesis, designing investigations, acquiring data, organizing data, analyzing investigations, concluding and decision making	High School (Grade 9)	Multiple choice	30 items No time mentioned
Abstract Operations Period Skills Test	Developed by: Demirbaş and Ertuğrul (2012)	Hypothetical thinking, probabilistic thinking, combinatorial thinking, correlational thinking, proportional thinking, identifying and defining variables	Primary School (Grade 4) Secondary School	Open ended	12 items No time mentioned
Test of Scientific Process Skills in Multiple Format	Developed by: Karslı and Ayas (2013)	Multiple choice: observing, measuring, classifying, predicting, determining variables, changing and controlling, stating hypothesis, interpreting data, drawing conclusion, designing and caring experiments Open-ended part: classifying, data recording, using data and modeling (graphic drawing), controlling variables and stating hypothesis, designing experiments	Undergraduate (Elementary science education)	Multiple choice Open ended	36 items 75-90 minutes
Science Process Skills Test	Developed by: Şardağ and Kocakülah (2016)	Observing, comparison, classifying, deduction, prediction, determination of variables, generating hypothesis, experimental design, controlling and manipulating variables, operational description, measurement, data recording, data processing, modeling, interpretation, deduction	Secondary School (Grade 8)	Multiple choice Open ended	Module I: 14 items Module II: 14 items 40 minutes
Science Process Skills Test	Developed by: Kurnaz and Kutlu (2016)	Observing, classifying and ranking, measuring, inquiring, hypothesizing, predicting, conducting and planning experiment, interpreting results, explaining the results	Primary School (Grade 4) Secondary School (Grade 5)	Multiple choice	39 items No time mentioned

interdisciplinary approach (cognitive development psychology, behavioral decision making researches, philosophy) to measure individuals' ability to evaluate scientific findings. With this scale, the scientific reasoning ability can be measured by concepts such as "blind / double blind experiments, causality, confounding variables, construct validity, control group, ecological validity, history, maturation, random assignment to conditions, reliability and response bias", which are related with science and scientific research processes. Literature review is conducted by using the expressions of "reasoning", "judgment" and "scientific process skills" and the measurement tools that are being used in Turkey are examined. Those that could be accessed are presented in Table 1 according to their various characteristics.

When Table 1 is examined, it appears that there are tools to measure the scientific process skills, logical thinking and reasoning of the students at different levels of education. Two different tests that measure the logical thinking skills (Aksu, Berberoğlu, & Paykoc, 1991; Sezen & Bülbül, 2011) and a test that measures scientific process skills (Karslı & Ayas, 2013), which are applied to students at undergraduate level, are seen and both of them are not towards scientific reasoning skill. In other words, a test for measuring undergraduate students' scientific reasoning level could not be detected, in Turkey. The absence of such a test was seen as a major deficiency and it was aimed to adapt and determine the psychometric properties of the Scientific Reasoning Scale (SRS) developed by Drummond and Fischhoff (2017) to Turkish.

Method

In this section, information on the steps followed during the process of adaptation of SRS to Turkish and the determination of psychometric properties are provided. In this context, processes for adaptation process, study groups, data collection tools and data analysis are presented respectively.

Adaptation Procedure

In the adaptation process of SRS, initially, Caitlin Drummond was contacted and her written consent was obtained. Then, the original form of the scale was translated into Turkish by three lecturers from English Language Teaching and English Language and Literature departments. These translations were comparatively examined by the researchers and the expressions that were considered as most appropriate for the items in the original form were gathered. During these examinations, changes were made such as using "Turkey" instead of "America" and transferring "wrong or right?" in item stems to the test instruction. The form obtained in this way was provided with the original form to a different faculty member in the Department of English Language and Literature who was not involved in the previous translation group and was checked to see if it was appropriate.

Turkish form was provided to a total of thirteen students at different levels; three of them at undergraduate, six of them at graduate and four of them at doctoral degrees along with an additional form requesting them to express what they understand from each item and the reason of each response they provide. In addition, interviews were conducted with three undergraduate students who did not take scientific research methods course, four undergraduate students who were enrolled in this course and five graduate students in order to examine the clarity of each item. As a result of these examinations, it was determined that the eighth item was understood by the participants in different ways. Some expressions regarding what students understand from this item and the reason of their responses are indicated in Table 2.

Table 2
Sample Expressions for Item Eight in Preliminary Examination

Status in question	Answer of student	Reason provided by student
• As the experiment progresses, the blue dot should flash faster.	False	• Instead of flashing of blue dot, something else should have been added in second stage.
• As the experiment progresses, the blue dot should flash faster.	False	• Rapid flashing of the blue dot reveals the need for individuals to make fast and instant decisions. This will increase the rate of mistake.
• Subjects are requested to push the button when the blue dot is flashing. Participants make more mistakes as the experiment progresses and the flashing speed has an impact on this.	True	• If the dot is flashing fast, the participants may have made more mistakes. The speed of dot has affected the error rate.
• The response time of subjects to the flashing of blue dot is investigated.	False	• Does the faster flashing of the blue dot mean that it appears briefly on the screen, or does it mean that the time between the two flashes is shortened? I am not sure. Whatever it is, it cannot be correct as the progress of experiment and the acceleration of blue dot are irrelevant.
• They are requested to press the button according to the blue dot on the computer screen.	False	• The reason of making more mistakes might be related with the problems of subjects. It might be the faster flashing.

When Table 2 is examined, it is understood that students generally understand the situation expressed in the item correctly but the reasons for their answers are insufficient or not related to the given situation. It has been considered that this problem in this item might occur again when implemented on larger groups and therefore an alternative item is added to the test in order to measure the same concept. These examinations and the Turkish form obtained after this addition were provided to a specialist who did not take part in the previous translation stages and the translation into English was conducted. The original form and the translated form were examined by both researchers and Caitlin Drummond. Drummond proposed partial correction for an item and stated that the added item could be substituted for the other. Following these transactions, linguistic equivalence study is initiated.

Study Groups

In the linguistic equivalence study, 35 lecturers who completed / attended graduate education and worked at different universities were included. The reason for the inclusion of this group in the linguistic equivalence study is the assumption that they have a certain level English language knowledge to answer both forms and that they

Table 3
Distribution of Study Groups in Different Stages of Research According to University, Gender, Department and Grade Levels

Stage‡		University and gender	Department / Grade	1	2	3	4	Total
Structural validity (n=429)		HKU: 239, Dokuz Eylül: 17, GAUN: 56, GU: 132, Female: 387, Male: 59	Elementary science education	-	-	17	-	17
			Elementary mathematics education	-	31	65	3	99
			Early childhood education	24	-	129	-	153
			Primary school education	25	-	122	-	147
			Turkish education	-	-	30	-	30
			Total	49	31	363	3	446
Criterion validity	CCTDI (n=140)	HKU: 19, Dokuz Eylül: 18, Ege: 9, GAUN: 12, GU: 147 Female: 170, Male: 35	Elementary mathematics education	-	-	10	7	17
			Primary school mathematics	-	12	-	1	13
			Early childhood education	8	-	-	28	36
			Psychological counseling and guidance	-	-	5	56	61
			Primary school education	9	11	11	47	78
			Total	17	23	26	139	205
	LCTSR (n=65)	GU: 57, Ege: 8 Female: 52, Male: 13	Elementary science education	-	-	2	55	57
			Primary school education	-	8	-	-	8
			Total	-	8	2	55	65
Comparison of groups (n=153)		HKÜ: 55, Dokuz Eylül: 16, Ege: 9, GAUN: 16, Gazi: 25, GU: 35 Female: 122, Male: 34	Elementary science education	-	-	16	-	16
			Elementary mathematics education	-	15	-	36	51
			Primary school education	7	10	25	-	42
			Early childhood education	12	-	-	-	12
			Psychological counseling and guidance -Graduate	-	-	-	-	23
			Psychological counseling and guidance -Dr	-	-	-	-	12
			Total	19	25	41	36	156
Pretest - Posttest		HKU: 41 Female: 37, Male: 4	Early childhood education	-	-	41	-	41
Test- retest		HKU: 34 Female: 34	Primary school education	-	-	18	-	18
			Psychological counseling and guidance	-	-	-	16	16
			Total	-	-	18	16	34
Grand Total								947

‡ The numbers in parentheses indicate the number of people remaining after listwise deletion and removing outliers
CCTDI: California Critical Thinking Disposition Inventory, LCTSR*: Lawson's Classroom Test of Science Reasoning*
HKU: Hasan Kalyoncu University, GAUN: Gaziantep University, GU: Giresun University

will not face any problems due to the fact that they are familiar with the concepts in the test. Cohen, Manion, and Morrison (2007) describe the data collection process as the appropriate sampling method, starting with the closest individuals, and including the accessible and available individuals at the time of the study. Similarly, Fraenkel, Wallen, and Hyun (2012) refer to the appropriate sample as the group of individuals eligible for research. The characteristics of the group studied on should be explained in detail and the research should be repeated on similar samples. Education faculty students included in this study were determined by appropriate sampling method and in the stages where the psychometric properties of SRS were examined, it was attempted to include different student groups as much as possible. Information on the study groups involved in these stages is given in Table 3.

As seen in Table 3, a total of 947 students were accessed, including 802 female and 145 male students studying at six different universities at different stages of their education.

Data Collection Tools

Scientific Reasoning Scale (SRS). The Scientific Reasoning Scale developed by Drummond and Fischhoff (2017) is the product of a comprehensive study conducted under a program supported by the National Science Foundation. It was developed specifically to determine whether individuals reject scientific evidence, or to what extent they can discern the bias caused by a false or incomplete presentation of a finding and SRS basically measures the ability to evaluate the quality of scientific evidence. Respondents are required to read the phrases on the scenarios prepared in the topics of blind / double-blind experimental studies, causality, confounding variables, construct validity, control group, ecological validity, history, maturation, random assignment to experimental conditions, reliability and response bias and they are asked to indicate whether it is true or false.

The final form of SRS was obtained as a result of different stages in which quantitative and qualitative examinations were carried out with a cyclical approach. In the first stage of data collection, 401 subjects ranging in ages of 18 -55, 20 items including attrition, measurement error and statistical power etc. were applied.

Correlations with Cognitive Reflection Test scores was .36; with numeracy measure was .28; for two different scientific literacy tests (TFKSS and USIS) were for .39 and .36, respectively and for those with undergraduate degrees, and for the number of science classes taken was .29 and significant ($p < .01$). SRS scores showing a relatively weak correlation with age (r=.14) and do not differ according to gender. Weak or insignificant correlations have been achieved with beliefs about global warming, vaccinations, and genetically modified foods, big bang and human evolution.

The factor loadings of the items in SRS which have a single factor structure are between .39 and .63; and their discriminations range from .43 to .55. The item difficulties calculated according to different studies are between .35 - .76 and .45 - .77. Confirmatory factor analysis results for the one-factor model indicate that data-model fit is approved ($\chi 2/sd$ = 3.09, RMSEA = .08, SRMR = .05, CFI = .90). The internal consistency coefficient of the test is .71.

It has been taken into consideration that the logic is one of the tools of scientific thinking (Özlem, 2014; Yıldırım, 2005) and that the individual's analysis, interpretation, questioning, explanation, evaluation and reflection of his own reasoning processes are in the center of critical thinking (Facione, 1990). It is considered appropriate to include the CCTDI and LCTSR in the examination of the criterion related validity of SRS.

California Critical Thinking Disposition Inventory (CCTDI). In the original form of the CCTDI, which emerged as a result of the Delphi project organized by the American Philosophical Society, there are 73 items under seven factors. The adaptation of the scale to Turkish has been carried out by Kökdemir (2003). As a result of the adaptation study, a single total score can be calculated from the scale consisting of six factors, namely analyticity, open-mindedness, inquisitiveness, self-confidence, truth seeking and systematicity, which are not very different from the original scale and reaching the 51-item form. It is stated that the total score from the CCTDI can also be used for the validity of educational programs designed to improve the tendency of critical thinking or skill (Kökdemir, 2003). The scale contains six response categories ranging from "Totally disagree" to "Totally agree". According to the exploratory factor analysis results, the factor loadings of the items ranged from .32 to .74. Six factors account for 36.13% of the total variance, and the item discriminations ranged from .20 to .50. The Cronbach α coefficients are between .61 and .78 for six factors and .88 for the whole scale. In this study, total scores of the CCTDI were used to examine criterion-related validity of SRS and Cronbach α coefficient for the whole scale is .67.

Lawson's Classroom Test of Science Reasoning (LCTSR). The test, which was developed by Lawson (1978) on the basis of cognitive development theory and then revised by Lawson et al. (2000), was adapted to Turkish by Yüzüak (2012). LCTSR consists of 24 multiple choice items that measure logical thinking and measure within the context of conservation of mass and volume, proportional thinking, control of variables, probabilistic thinking, correlational thinking, and hypothetical thinking skills. The responses given to the item pairs are considered together when scoring. In order to get one point from a question, the second question, which explains the previous one needs to be answered correctly. In the adaptation study conducted by

Yüzüak (2012), it is stated that item difficulties vary between .16 - 1.00 and their discriminations vary between .00 - .58. The relationship between the subscale scores and the total score ranged from .20 to .63, and split half reliability was .72. The adaptation study was conducted for the purpose of examining the criterion validity of LCTSR which was carried out on high school students considering that SRS showed similarity both in terms of content and scoring process. In this study, the internal consistency of LCTSR calculated with the KR-20 formula is .55.

Data Analysis

The correlations of item scores between SRS's original and Turkish forms were examined with tetrachrotic and the total scores with the Pearson correlation coefficient. Principal component analysis (PCA) and confirmatory factor analysis (CFA) were performed in the scope of construct validity. Criterion - related validity with CCTDI and LCTSR and test - retest reliability were investigated with Pearson correlation coefficient. Undergraduate and graduate level students' SRS scores were compared by independent samples t test. The scores obtained from the applications of SRS at the beginning of spring semester of 2017-2018 before and after attending to the course of scientific research methods were compared with the t test for related samples. The effect sizes were interpreted over .2, .5 and .8, respectively, for small, medium and wide effect according to Cohen d (Büyüköztürk, 2018). For the calculation of the internal consistency coefficients, Cronbach α for the CCTDI and KR-20 for the LCTSR. The discrimination of SRS items was examined by biserial correlation coefficient.

Prior to statistical analysis, we examined the rate and pattern of missing data, whether there were outliers in the data sets, and whether there was a significant violation of the normality of the distribution. In the datasets, missing data ratios were between 0% - 0.8% showing missing completely at random pattern. Listwise deletion was used and these rows were not included in the analyzes (Akbaş, 2017). The rows with standard scores standard beyond the range of [-3, + 3] has been discarded. The significance level in analyzes was .05.

Findings

Findings Related to SRS's Linguistic Equivalence

Within the scope of the linguistic equivalence study, a group of 35 people consisting of the lecturers and research assistants working in different institutions were selected. English and Turkish forms were applied with 12-15 days intervals. In this phase, 15 individuals have first taken the Turkish and then the English form, and 20 individuals have first taken the English form and then the Turkish form. Correlation coefficients calculated for each item pairs and the total scores are given in Table 4.

Table 4
Correlation Coefficients Between the Items and the Total Scores of English and Turkish Forms

Item	Correlation	Item	Correlation
1	.70	7	.95
2	.90	8	1.00
3	.72	9	.90
4	.78	10	.94
5	1.00	11	.73
6	.95	Total Score	.91

In Table 4, it is seen that the correlation coefficients between item pairs in Turkish and English forms are between .70 - 1.00. Regarding the translation process and the correlation coefficients obtained, it is understood that the linguistic equivalent of SRS is approved.

Findings Related to SRS's Factor Structure

Güngör (2016) states that PCA is used when there is no prior knowledge and CFA is used when testing an existing theory (e.g. factor structure). PCA is appropriate for the test development while CFA is appropriate for the adaptation process. Çokluk, Şekercioğlu, and Büyüköztürk (2014) indicate that CFA is an analysis in which a previously defined and restricted structure is tested as a model. It is known that SRS consists of a single factor (Drummond & Fischhoff, 2017). Nevertheless, PCA based on the tetrachoric correlation matrix was carried out to investigate the changes in the factor structure which came up from alternative item. So, two PCA's and CFA's were performed, one for the test which includes original eighth item and one for the test with the alternative of this item. PCA results are given at Table 5.

Table 5
Results of Principal Component Analysis for the Form Containing the Original Item and the Alternative

Item	Factor loadings (Model 1)	Factor loadings (Model 2)
1	.68	**.67**
2	.63	**.63**
3	.68	**.68**
4	.66	**.64**
5	.53	**.53**
6	.58	**.59**
7	.61	**.61**
8	.10	**.57**
9	.57	**.56**
10	.57	**.58**
11	.60	**.63**
Eigenvalues$^{\beta}$	3.71 (1.26) 1.14 (1.19) 1.02 (1.13)	**4.04 (1.26)** **1.04 (1.19)**
Extracted variance	33.72 %	**37.76 %**
Bartlett	429.3 ($p < .00$)	**496.6 ($p < .00$)**
KMO	.80	**.82**

β Eigenvalues given in parentheses were obtained by parallel analysis.

In Table 5, Bartlett test of sphericity for both models were significant (p <.00) and the KMO values are found to be .80 and .82, respectively. Comparing the eigenvalues obtained for different data sets with the eigenvalues obtained by parallel analysis (Watkins, 2000), it can be understood that one can speak of unidimensional structure for both sets of data. The extracted variance model 1 is 33.72% while for model 2 it is 37.76%. For model 1, the factor loading of the eighth item is .10, while the factor loadings of the other items are in the range of .53 - .68. In model 2 all factor loadings are in the range of .53 - .68. Two CFA's were performed and path diagrams with standard coefficients are given in Figure 1 and Figure 2.

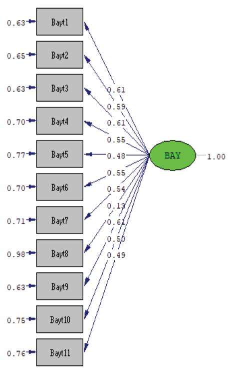

Figure 1. Path diagram (Model 1).

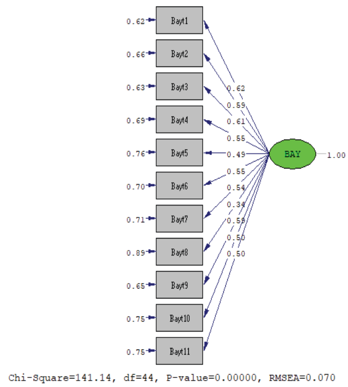

Figure 2. Path diagram (Model 2).

Figure 1 indicates that the factor loading of the eighth item is lower for model 1 (λ_8=.13, $p > .01$) than the model 2 (λ_8=.34, $p < .01$). The fit indexes obtained for both models and the recommendations according to the literature (Browne & Cudeck, 1993; Hu & Bentler, 1999; Kline, 2005) are given in Table 6.

Table 6
Model Fit Indexes Obtained as a CFA and Recommendations

Index	Model 1	Model 2	Recommendation
χ^2/sd	170.85/44=3.88	**141.14/44=3.21**	Perfect ≤ 3 ≤ Good ≤ 5
RMSEA	.08	**.07**	Perfect ≤ .05 ≤ Good ≤ .08
GFI	.93	**.95**	Perfect ≤ .95 ≤ Good ≤ .90
AGFI	.90	**.92**	Perfect ≤ .95 ≤ Good ≤ .90
CFI	.93	**.95**	Perfect ≤ .95 ≤ Good ≤ .90
NFI	.91	**.93**	Perfect ≤ .95 ≤ Good ≤ .90
NNFI	.91	**.94**	Perfect ≤ .95 ≤ Good ≤ .90
RMR	.06	**.05**	Perfect ≤ .05 ≤ Good ≤ .08
SRMR	.06	**.05**	Perfect ≤ .05 ≤ Good ≤ .08

When the values given in Table 6 are examined, it is seen that both models fit and the values obtained for model 2 are better. The PCA and CFA results and the difficulties in understanding the eighth item were taken into account together and the following stages were carried out on the form containing the alternative item.

Findings Related to SRS's Criterion Validity

The correlations between SRS scores and the CCTDI and LCTSR scores were examined within the context of the criterion validity. As a result, correlation between SRS and CCTDI total scores were .32 ($p < .01$); LCTSR total scores were .46 ($p < .01$). Büyüköztürk (2018) suggests that the correlation coefficients between .30 - .70 can be interpreted as indicating a moderate relation. According to this, there is a positive and moderate relation between the total scores of SRS and the total scores of the CCTDI and LCTSR within the scope of criterion validity.

Comparison of Groups Trained at Undergraduate and Graduate Levels

SRS total scores were compared with independent samples t-test over the mean scores of students studying at undergraduate and graduate level and results are given at Table 7.

Table 7
T Test Results on the Comparison of SRS Mean Scores According to Education Level

Education Level	N	Mean	Standard Deviation	df	t	Cohen d
Undergraduate	169	5.36	1.64	202	3.67**	.68
Graduate	35	6.49	1.67			

**$p < .01$.

Table 7 indicates that the mean of students at graduate level was significantly higher than the mean of students at the undergraduate level ($p < .01$), and a moderate effect size was observed (Cohen d = .68).

Although there are limitations about stating a control group, random assignment to conditions and controlling for other variables, it is foreseen that scientific research methods course carried out during a semester will lead to a significant increase in the SRS scores. SRS was applied as pre-test in the first week and as post-test in the last week during a scientific research methods course, in the spring term of 2017 – 2018. Scores were compared by using independent samples t-test and the results are given in Table 8.

Table 8
T Test Results on SRS's Application Before and After Scientific Research Methods Course

Application	N	Mean	Standard Deviation	df	t	Cohen d
Pre-test (first week)	41	4.63	1.88	40	3.57**	.56
Post-test (last week)	41	5.88	1.62			

**$p < .01$.

In Table 8, it is seen that the mean of SRS scores is 4.63 before the course and 5.88 at the end of the course. Scores showed a significant increase in favor of the post-test ($p < .01$) and a moderate effect size is apparent (Cohen d = .56). This finding supports the prediction that the level of scientific reasoning will increase at the end of the scientific research course.

Reliability of SRS

SRS's reliability is examined by test-retest, internal consistency (KR-20) and split-half methods. The correlation coefficient between total scores obtained from SRS's 14 - 21 day interval applications is .78, the KR-20 coefficient calculated based on the data set in which the structure validity is examined is .70; split-half reliability is .68.

Item Analysis

Difficulty and discrimination indices of SRS items were calculated using the dataset which DFA was performed are given in Table 9. As this study aimed to adapt SRS to Turkish, items provided here are in Turkish. Original items can be reached in Drummond and Fischhoff (2017).

Table 9
Item Statistics

Items[ϕ]	Difficulty	Discrimination
1. Bir lezzet testinde araştırmacı, A markalı kahveyi beyaz bantlı kupaya, B markalı kahveyi aynı tip siyah bantlı kupaya koymuştur. Laboratuvar asistanı kupaları katılımcılara verirken araştırmacı da katılımcıların yüz ifadelerini izler. Laboratuvar asistanı hangi kupada hangi kahvenin bulunduğunu bilmemelidir. (*true*)	.32	.71
2. Bir araştırmacı Türkiye'de daha geniş ormanlık alana sahip bölgelerde nesli tükenmekte olan hayvan sayısının daha az olduğunu görmüştür. Bu veriler, Türkiye'de ormanlık alanların genişliğini artırmanın nesli tükenmekte olan hayvan sayısını azaltacağını göstermektedir. (*false*)	.31	.67
3. Bir araştırmacı deneklerin bir kısmını yüksek sesli radyo yayını yapılan ve soğuk bir odaya; bir kısmını da radyo yayını yapılmayan sıcak bir odaya koyar ve verilen yap-bozu yapmalarını ister. Radyo yayını yapılmayan ve sıcak odadaki denekler yap-bozu daha hızlı yaparlar. Araştırmacı, diğer odadaki katılımcıların yap-bozu daha yavaş yapmalarının nedeninin radyo yayını olup olmadığını söyleyemez (*true*)	.34	.70
4. Bir eğitim araştırmacısı matematikte yüksek performans sergileyen öğrencilerin genel matematik yeteneğini ölçmek istemektedir. Bütün öğrenciler geometri ve temel matematiğe giriş derslerini almışlardır. Araştırmacı öğrencilerin genel matematik yeteneğini geometri testi kullanarak ölçebilir. (*false*)	.30	.69
5. İki araştırmacı sivilce sorunu olan ergenler üzeirnde bir sivilce kreminin etkisini test etmektedir. Bu araştırmacılardan biri kremi çalışmaya katılan tüm ergenlere vermek istemektedir. Diğeri ise grubun bir yarısına sivilce kremini uygularken diğer yarısına içerisinde sivilce önleyici madde olmayan bir başka krem uygulamayı istemektedir. Her iki yöntem de kremi test etmede eşit derecede etkilidir. (false)	.52	.61
6. Bir araştırmacı bir grup deneğe rekabete dayalı bir oyun oynatacaktır. Her deneğin amacı jeton alıp satarak para kazanmaktır. Deneklere bu deneye katılmaları karşılığında sabit bir ücret ödenmektedir. Araştırmacı deneydeki davranışların gerçek hayatta alım satım davranışını yansıttığını güvenle söyleyebilir. (*false*)	.53	.65
7. Rastgele seçilmiş bir grup bireyin A hastalığı hakkındaki görüşleri, hastalıkla ilgili yapılan altı aylık bir medya kampanyasından önce ve bu kampanyadan sonar alınmıştır. Anket sonuçlarına göre, katılımcıların hastalık hakkındaki bilgisi medya kampanyasından sonra artmıştır. Kampanya, hastalık hakkındaki bilgileri arttırmamış olabilir. (*true*)	.48	.57

Table 9
Item Statistics

Items^ϕ^	Difficulty	Discrimination
8. Bir hastanenin çocuk hastalıkları bölümüne astım şikayeti ile gelen çocuklara özel bir tedavi programı uygulanmaktadır. Bu programın uygulandığı çocukların, ilerleyen zamanlarda benzer şikayetlerle bu servise tekrar gelme sıklıklarında azalma görülmüştür. Uygulanan özel tedavi yöntemi astım hastalığı üzerinde iyileştirici etkiye sahiptir. (*false*)	.44	.50
9. Araştırmacılar bir beslenme programının çocukların kilo vermesine yardımcı olup olmadığını öğrenmek istemektedirler. Çocuklar deney ve control grubu olarak ikiye ayrılırlar. Araştırmacılar fazla kilolu çocukları deney grubuna koymalıdırlar. (*false*)	.50	.68
10. Bir araştırmacı sıvılardaki yüzey gerilimini ölçmek için yeni bir metot geliştirir. Bu metot eski metota göre daha tutarlıdır. Bu durum, yeni metodun eskisine göre daha isabetli sonuçlar verdiği anlamına gelir. (*false*)	.39	.62
11. İki araştırmacı, tüketicilerin müşteri hizmetlerinden duydukları memnuniyeti ölçmek için birer anket geliştirir. Araştırmacılar müşterilerinden, beş dereceli bir ölçek üzerinde "Müşteri hizmetlerinden memnunum" ifadesine katılma derecelerini belirtmelerini ister. A araştırmacısı dereceleri 1: Kesinlikle katılıyorum" ve 5: "Kesinlikle katılmıyorum" şeklinde; B araştırmacısı ise 1: "Memnun olmadığımı söyleyemem" ve 5: "Son derece memnuniyetsizim" şeklinde belirler. Bu dereceler tüketicilerin müşteri hizmetlerinden duyduğu memnuniyeti ölçmede birbirine eşdeğerdir. (*false*)	.30	.61

ϕ Correct answers are indicated in parentheses.

Table 9 indicates that item difficulties are between .30 and .53; while the discriminations vary between .50 and .71. Average difficulty of items is .40 and average discrimination is .64 According to this, SRS can be defined as a test with moderate difficulty and high discrimination.

Conclusion and Discussion

In this study, SRS, which is developed by Drummond and Fischhoff (2017) is adapted to Turkish. According to investigations conducted for the linguistic equivalence, it was determined that the eighth item related to the concept of maturation for the experimental researches was not sufficiently understood and a new item was added to the test instead of it. The high correlations between item pairs and total scores in Turkish and English forms have shown that linguistic equivalence is provided.

The results of PCA and CFA indicated that SRS had a single factor structure. This single factor accounts for approximately 38% of the variance. Büyüköztürk (2018) states that the extracted variance ratio of 30% or more by a single-factor may be sufficient. Accordingly, it can be said that the rate of variance extracted by SRS is sufficient. Moderate and positive correlations with LCTSR and CCTDI total scores within the scope of the criterion validity, supports the conceptual relationship between scientific reasoning and logical thinking and critical thinking. Comparing the undergraduate students with graduate students in terms of SRS scores; the scores after taking scientific research methods course is higher than the scores before this course.

For the tests used in education and psychology, it is recommended that the reliability should be at least .70 (Nunnaly, 1978). Internal consistency coefficients for LCTSR and CCTDI, which are included in the criterion-related validity, were not at the expected level (.55 and .67 respectively) In the study which LCTSR adapted to Turkish by Yüzüak (2012) split half reliability of test was .67. And it was reported that internal consistency coefficients for the dimensions and whole scale of CCTDI changed between .61 - .88 (Kökdemir, 2003). It is considered that the reliability coefficients obtained for these scales applied to smaller groups according to the mentioned adaptation studies are acceptable.

Reliability coefficients obtained by different methods for SRS were found to vary between .68 and .78. It can be said that these reliability values are acceptable when it is taken into consideration that some items in SRS have low item difficulty indices (i.e. p_2=.31, p_4=.30 and p_{11}=.30). KR-20 coefficient is calculated based on item difficulties (Baykul, 2000). Because, the contribution of these items to the observed score variance is relatively low.

One test to measure the logical thinking ability of undergraduate students in Turkey based on Piaget's theory of cognitive development is Group Assessment of Logical Thinking (GALT). In this test consevation, length/volume, proportional reasoning, controlling variables, combinatorial reasoning, probabilistic reasoning, correlational reasoning are measured (Aksu et al., 1990). Another test is the Logical Thinking Test (LTT) developed by Sezen and Bülbül (2011). This test measures the ability to defining and controlling variables, associating, calculating probability, interpreting graphics, transforming numerical expressions into graphs. Test of Scientific Process Skills in Multiple Format (TSPMF) is developed by Karslı and Ayas (2013) and allows measuring such traits as observing, measuring, classifying, predicting and manipulating variables. These tests, while not covering scientific reasoning skills also have limitations in terms of the applicable groups. For example, LTT can be applied to students who are studying in undergraduate programs of mathematics education, while TSPMF can be applied to students who are studying in undergraduate programs of elementary science education. SRS offers measurements of blind/double-blind experiments, causality, confounding variables, construct validity, control group, ecological validity, history, maturation, random assignment to conditions, reliability, response bias, and is separated from these tests in terms of its applicability to various programs of the education faculties.

It can be argued that SRS, which consists of true / false is objectively scored and practical as it takes about ten minutes to answer. Wooley et al. (2018) point out that the students of the undergraduate level can have trouble following the scientific reasoning process, while Kuhn, Ramsey, and Arvidsson (2015) state that the scientific thinking skill increases with the level of education. The fact that the SRS scores of the students who are studying at the graduate level are significantly higher than those of the undergraduate students is in agreement with the findings of these researches. When the findings are evaluated as a whole, it can be said that SRS has suitable psychometric properties to be used in the researches that aim to determine the scientific reasoning skills of undergraduate and graduate students.

SRS provides information on eleven different concepts such as causality, confounding variable, and reaction bias. In the development process, some items that measure concepts such as attrition, measurement error, selection bias, statistical power have been excluded because of a variety of reasons i.e. low factor loading or misunderstanding (Drummond & Fischhoff, 2017). In other words, SRS does not contain some components that may be included in scientific reasoning. Depending on this situation, a reasoning test that measures these concepts is needed as a complement to SRS.

References

Adey, P., & Csapo, B. (2012). Developing and assessing scientific reasoning. In B. Csapó & G. Szabó (Eds.), *Framework for diagnostic assessment of science* (pp. 17–53). Budapest: Nemzeti Tankönyvkiadó.

Akbaş, U. (2017). Examination of the effects of different missing data techniques on item parameters obtained by CTT and IRT. *International Online Journal of Educational Sciences*, *9*(3), 603–616.

Aksu, M., Berberoğlu, G., & Paykoç, F. (1991). *Mantıksal düşünmenin belli değişkenlere göre incelenmesi*. Paper presented at Eğitimde arayışlar I. sempozyumu, İstanbul, Turkey.

Aydoğdu, B., Tatar, N., Yıldız, E., & Buldur, S. (2012). İlköğretim Öğrencilerine Yönelik Bilimsel Süreç Becerileri Ölçeğinin geliştirilmesi [The science process skills scale development for elementary school students]. *Journal of Theoretical Educational Science*, *5*(3), 292–311.

Bady, R. J. (1979). Students' understanding of logic hypothesis testing. *Journal of Research in Science Teaching, 16*(1), 61–65.

Barbey, K. A., & Barsalou, L. W. (2009). Reasoning and problem solving: models. *Encyclopedia of Neuroscience, 8*, 35–43.

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik Test Teorisi ve uygulama* [Measurement in education and psychology: Classical Test Theory and application]. Ankara, Turkey: ÖSYM Yayınları.

Beyer, B. (1988). Developing a scope and sequence for thinking instruction. *Educational Leadership, 45*(7), 26–30.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Research & Methods*, *21*, 230–258.

Butterworth, J., & Thwaites, G. (2013). *Thinking skills-critical thinking and problem solving*. Edinburgh: Cambridge University Press.

Büyüköztürk, Ş. (2018). *Sosyal bilimler için veri analizi el kitabı: istatistik, araştırma deseni, SPSS uygulamaları ve yorum* [Data analysis handbook for social sciences: statistics, research design, SPSS applications and interpretation]. Ankara, Turkey: Pegem Akademi.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York: Routledge.

Cüceloğlu, D. (2005). *İyi düşün doğru karar ver*. İstanbul, Turkey: Remzi Kitapevi.

Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2016). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* [Multivariate statistics for social sciences: SPSS and LISREL applications]. Ankara, Turkey: Pegem Akademi.

Demirbaş, M., & Ertuğrul, N. (2012). Fen ve teknoloji dersine ilişkin Piaget'in soyut işlemler döneminde kazanılması beklenen becerilerin gerçekleşme durumunun incelenmesi [An investigation into the realization of skills in the science and technology lessons expected to be acquired in Piaget's abstract operations stage]. *Kalem Eğitim ve İnsan Bilimleri Dergisi*, *2*(2), 123–164.

Dewey, J. (1910). *How we think?* Washington DC: Heath & CO. Publishers.

Drummond, C., & Fischhoff, B. (2017). Development and validation of the scientific reasoning scale. *Journal of Behavioral Decision Making, 30*, 26–38.

Dunbar, K., & Fugelsang, J. (2004). Scientific thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp.1–9). Edinburgh: Cambridge University Press.

Facione, P. A. (1990). *Critical thinking: A statement of experiment consensus for purposes of educational assessment and instruction* (*The Delphi Report*). Project number: ED315423. Retrieved from https://philarchive.org/archive/FACCTA

Feyzioğlu, B., Demirdağ, B., Akyıldız, M., & Altun, E. (2012). Ortaöğretim öğrencilerine yönelik bilimsel süreç becerileri testi geliştirilmesi: geçerlik ve güvenirlik çalışması [Developing a science process skills test for secondary students: validity and reliability study]. *Kuram ve Uygulamada Eğitim Bilimleri, 12*(3), 1887–1906.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. New York: McGraw Hill.

Geban, Ö., Aşkar, P., & Özkan, İ. (1992). Effects of computer simulations and problem-solving approaches on high school students. *Journal of Educational Research*, *86*(1) 5–10.

Güneş, F. (2012). Öğrencilerin düşünme becerilerini geliştirme [Improving the thinking skills of students]. *Türklük Bilimi Araştırmaları Dergisi*, *32*, 127–146.

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

Han, J. (2013). *Scientific reasoning: research, development, and assessment* (Doctoral dissertation). Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=osu1366204433

Hogan, K., & Fisherkeller, J. (2005). Dialogue as data: Assessing students' scientific reasoning with interactive protocols. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 95–127). Burlington, MA: Elsevier Academic Press.

Holyoak, K. T., & Morrison, R. G. (2004). Thinking and reasoning: A reader's guide. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 1–9). Edinburgh: Cambridge University Press.

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

Karslı, F., & Ayas, A. (2013). Fen ve teknoloji dersi öğretmen adaylarının bilimsel süreç becerilerinin ölçülmesine ilişkin bir test geliştirme çalışması [A study on the development of a test about measurement of science and technology teachers' science process skills]. *Türk Fen Eğitimi Dergisi*, *10*(1), 66–84.

Kind, P., & Osborne, J. (2016). Styles of scientific reasoning: a cultural rationale for science education? *Science Education*, *101*(1), 8–31.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. NY: Guilford Publications, Inc.

Kökdemir, D. (2003). *Belirsizlik durumlarında karar verme* [Decision making and problem solving under uncertainty] (Doctoral dissertation, Ankara University, Ankara, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Kuhn, D., Ramsey, S., & Arvidsson, T. S. (2015). Developing multivariate thinkers. *Cognitive Development, 35*, 92–110.

Kurnaz, F. B., & Kutlu, Ö. (2016). İlkokul 4. sınıf için hazırlanan bilimsel süreç becerileri programının etkililiğinin belirlenmesi [Determining the effectiveness of science process skills program prepared for elementary school grade 4]. *İlköğretim Online*, *15*(2), 529–547.

Lawson, A. E. (1978). The development and validation of classroom test of formal reasoning. *Journal of Research in Science Teaching, 15*(1), 11–24.

Lawson, E. A., Clark, B., Meldrum-Cramer, E., Falconer, A. K., Sequist, M. J., & Kwon, Y. (2000). Development of scientific reasoning in college biology: do two levels of general hypothesis-testing skills exist? *Journal of Research in Science Teaching, 37*(1), 81–101.

Lipman, M. (2003). *Thinking in education*. United Kingdom: Cambridge Press.

McGuinness, C. (2005). Teaching thinking: Theory and Practice. *British Journal of Educational Psychology Monograph Series II: Pedagogy - Teaching for Learning*, *3*, 107–126.

Morgan, C. T. (2013). *Psikolojiye giriş* [Introduction to psychology] (S. Karakaş & R. Ekşi, Trans.). Konya, Turkey: Eğitim Yayınevi.

Nessel, D. D., & Graham, J. M. (2007). *Thinking strategies for student achievement: Improving learning across the curriculum, K-12*. California: Corwin Press.

Novia, N., & Riandi,R. (2017). The analysis of students scientific reasoning ability in solving the modified Lawson Classroom Test of scientific reasoning (MLCTSR) problems by applying the levels of inquiry. *Jurnal Pendidikan IPA Indonesia, 6*(1), 116–122.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Önal, İ. (2005). *İlköğretim fen bilgisi öğretiminde performans dayanaklı durum belirleme uygulaması üzerine bir çalışma* [A study about performance based assessment in primary science education]. (Master's thesis, Hacettepe University, Ankara, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Özden, Y. (2011). *Öğrenme ve öğretme* [Learning and teaching]. Ankara, Turkey: Pegem Akademi.

Özlem, D. (2014). *Mantık* [Logic]. İstanbul, Turkey: Notos Kitap Yayınevi.

Öztürk, N. (2008). *İlköğretim yedinci sınıf öğrencilerinin fen ve teknoloji dersinde bilimsel süreç becerilerini kazanma düzeyleri* [Primary seventh grade students' level of gaining science process skills in science and technology course] (Master's thesis, Osmangazi University, Eskişehir, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Paul, R., & Elder, L. (2014). *Crtical thinking: tools for taking charge of your professional and personal life.* New Jersey: Pearson Education, Inc.

Schauble, L. (1996). The development of scientific reasoning in knowledge rich contexts. *Developmental Psychology, 32*(1), 102–119.

Sezen, N., & Bülbül, A. (2011). A scale on logical thinking abilities. *Procedia Social and Behavioral Sciences, 15*, 2476–2480.

Siegel, H. (1999). What (good) are thinking dispositions? *Educational Theory*, *49*(2), 207–224.

Stuessy, C. S. L. (1984). *Correlates of scientific reasoning in adolescents: Experience, locus at control, age, field dependence-independence, rigidity/flexibility, IQ, and gender* (Doctoral dissertation, Ohio State University). Retrieved from https://eric.ed.gov/?id=ED244834

Şardağ, M., & Kocakülah, A. (2016). Sekizinci sınıf öğrencilerine yönelik bir bilimsel süreç becerileri testi geliştirme çalışması [Science process skills test development study towards the eighth grade students]. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, *31*, 1–32.

Tatar, N. (2006). *İlköğretim fen eğitiminde araştırmaya dayalı öğrenme yaklaşımının bilimsel süreç becerilerine, akademik başarıya ve tutuma etkisi* [The effect of inquiry-based learning approaches in the education of science in primary school on the science process skills, academic achievement and attitude] (Doctoral dissertation, Gazi University, Ankara, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Tobin, K. D., & Capie, W. (1981). Development and validation of a group test of logical thinking. *Educational and Psychological Measurement, 41*(2), 413–424.

Türk Dil Kurumu. (2018). *Türkçe sözlük* [Turkish dictionary]. Retrieved from http://www.tdk.gov.tr/

Türker, E. (2011). *Bilimsel süreç becerileri yaklaşımının model kullanarak uygulanmasının öğrencilerin başarılarına, bilimsel süreç becerilerinin gelişimine ve motivasyonlarına etkisi* [To investigate how scientific process skills approach based on model using affect the students? success, development of process skills and motivations] (Master's thesis, Karadeniz Teknik University, Trabzon, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Watkins, M. W. (2000). *Monte Carlo PCA for parallel analysis* [computer software]. State College, PA: Ed & Psych Associates.

Woolley, J. S., Deal, A. M., Green, J., Hathenbruck, F., Kurtz, S. A., Park, T. K. H., Pollock, S. V., … Jensen, J. L. (2018). Undergraduate students demonstrate common false scientific reasoning strategies. *Thinking Skills and Creativity, 27*, 101–113.

Yıldırım, C. (2005). *Bilim felsefesi* [Philosophy of science]. İstanbul: Remzi Kitapevi.

Yüzüak, A. V. (2012). *Lawson mantıksal düşünme testinin uyarlanması ve uygulanması* [Adaptation and application of Lawson's classroom test of science reasoning] (Master's thesis, Gazi University, Ankara, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school, *Developmental Review, 27*, 172–223.