A Meta-Analytic Reliability Generalization Study of the Computational Thinking Scale‡

Gülbin Kiyici,[1] Nurcan Kahraman[2]

1. *Manisa Celal Bayar University, Faculty of Education, Turkey*
2. *Bursa Uludag University, Faculty of Education, Turkey*

***Abstract**: This study aims to analyze the reliability generalization of the computational thinking scale. There are five dimensions of computational thinking: creativity, algorithmic thinking, cooperativity, critical thinking, and problem-solving. A Bonett transformation was used to standardize the reliability coefficient of Cronbach's alpha. A random-effects meta-analysis was conducted since the heterogeneity among the studies was high. Results supported the RG of the computational thinking scale and its sub-dimensions, which were calculated as 0.843 for general, 0.799 for creativity, 0.848 for algorithmic thinking, 0.863 for cooperativity, 0.799 for critical thinking, and 0.817 for problem-solving. Besides that, the moderator analysis was conducted for the sample type, test length, country, and language of the study. According to the findings, there were no significant moderator effects on the reliability estimation.*

Science Insights Education Frontiers 2022; 13(2):1859-1874.
Doi: 10.15354/sief.22.ma011

How to Cite: Kiyici, G., & Kahraman, N. (2022). A meta-analytic reliability generalization study of the computational thinking scale. Science Insights Education Frontiers, 13(2):1859-1874.

***Keywords:** Reliability Generalization, Meta-Analysis, Computational Thinking*

***About the Authors**: Gülbin Kiyici, PhD, Manisa Celal Bayar University, Faculty of Education, Turkey, E-mail: gulbin.kiyici@gmail.com*

Nurcan Kahraman, Bursa Uludag University, Faculty of Education, Turkey, PhD, E-mail: nurcankahraman@uludag.edu.tr

***Correspondence to**: Dr. Gülbin Kiyici at Manisa Celal Bayar University of Turkey.*

***Conflict of Interests**: None*

Introduction

Computational thinking (CT) can be defined as thinking processes that involve formulating problems, apart from programming skills (Clark, 2015; Denning, 2017). In other words, developing solutions to the problems via algorithms and steps are degrees of CT (Aho, 2012). For instance, individuals who have developed computational thinking focus on analyzing a problem, collecting data to solve the problem, and dividing the solution into sub-steps. These individuals are not concerned with the outcome of the problem, but with the solution process of the problem and how it applies to similar problems (Vaidyanathan, 2016). Developing the CT skills of individuals is not only useful to reach goals and find solutions; it also supports other skills such as thinking scientifically, doing arithmetic, reading, and writing (Grover & Pea, 2013; Papert, 1996; Wing, 2006). Hence, individuals with a high level of computational thinking are also expected to show high academic success in different fields (Lei et al., 2020). In the relevant literature, researchers generally prefer to explain the functions, skills, and related concepts of computational thinking rather than describing them. For example, the International Educational Technologies Community states that computational thinking includes the skills of creativity, algorithmic thinking, critical thinking, problem-solving, communication, and cooperation (ISTE, 2019). Therefore, it is not surprising that supporting computational thinking will also help to develop 21st-century skills such as problem-solving, logical reasoning, and analytical thinking (Hunt & Riley, 2014).

An important question about CT is how to measure it. The proposed measurement tools generally cope with this issue by focusing on different dimensions of CT. For example, Akyol (2019) developed a CT scale and assessed the following dimensions of CT: computational thinking, robotic coding and software, professional development, and career planning. In another scale developed by Tsai and his colleagues (2021), evaluation, abstraction, decomposition, generalization, and algorithmic thinking appear as sub-dimensions of CT (Tsai et al., 2021). Another measurement tool developed by Korkmaz et al. (2017) is the most frequently used scale to determine the level of CT in the relevant literature (Avcu & Ayverdi, 2020; Çakır & Yaman, 2018; Çevik et al., 2021; Karaçaltı & Korkmaz, 2018; Oluk et al., 2018; Özgür, 2020; Yalçın & İkinci, 2020). Hence, this study aims to conduct a reliability generalization (RG) of the CT scale devised by Korkmaz et al. (2017)'s CT scale.

Reliability Generalization

Reliability is a problem related to the measurement precision of test scores. To solve this problem, researchers record the number of errors of repeated measurements and determine how precisely the scale measures the specified quality (McDonald, 1999). It can be said that reliability is the degree to which the measurements are free from random error. In addition, reliability means the consistency of the scores of individuals

‡: *Supplementary materials are included for this article.*

participating in a test (Anastasi, 1976). Therefore, the reliability of the scores obtained from a measurement tool must be investigated at every turn (Crocker & Algina, 2008). Cronbach's alpha, a coefficient used in the literature to analyze and interpret reliability, can be dissimilar for each measurement even though the same scale is used because it is a sample-dependent coefficient. Interpreting the extent of observed effects requires an assessment of the reliability of the scores. Therefore, the reliability coefficients of the scores obtained from the scale for the analyzed data should be reported (Wilkinson, 1999). RG studies aim to explain the source of the variability in the reported reliability values in different applications of a scale through a meta-analysis study. According to Vacha-Haase (1998), RG can be used to estimate the pooled score reliability of the scale used in different studies. In other words, RG studies are carried out to determine the level and source of the difference between the reliability coefficient of the studies (Eser & Aksu, 2021; Liang et al., 2021; Olderbak et al., 2021; Opitz et al., 2020; Yoon et al., 2021).

The study aimed to examine the RG of the computational thinking scale (Korkmaz & Bai, 2019; Korkmaz & Özden, 2015; Korkmaz et al., 2017). It consists of 29 questions in total divided into five sub-dimensions: creativity (8), algorithmic thinking (6), cooperativity (4), critical thinking (5), and problem-solving (6). The questionnaire was developed for undergraduate students and then it was adapted for middle school students. The adapted version of the questionnaire consists of 22 items with similar sub-dimensions. There were two main reasons for choosing this scale in this study. Firstly, although CT is mostly associated with computer science, it is a skill that can be associated with all other fields. Besides that, the number of related studies has increased with the increasing importance of CT (Kıyıcı & Yamak, 2021). Hence, it is important to ascertain the reliability of the instrument, and secondly, the mentioned scale is frequently used to assess CT in the literature (e.g. Avcu & Ayverdi, 2020). Since the language of the instrument is Turkish, the reliability can be generalized to the Turkish sample. To sum up, the present study aims to answer the following research questions:

i. What is the overall reliability estimate of the CT scale?
ii. To what extent does the reliability estimate of the CT scale vary by year and language of the article?
iii. To what extent does the reliability estimate of the CT scale vary by country or level of the participants?

Method

This study aimed to calculate the pooled reliability coefficient of the computational thinking scale (Korkmaz & Bai, 2019; Korkmaz & Özden, 2015; Korkmaz et al., 2017). For this aim, the recommendations of the Cochrane Handbook for systematic reviews of primary research were followed (Higgins et al., 2003). The findings of the study were reported according to the PRISMA statement for reporting systematic reviews and meta-analyses of studies (Liberati et al., 2009).

Sample of Studies

To identify studies, online databases such as Web of Science, the Educational Resources Information Center (ERIC), and Google Scholar were searched by the research-

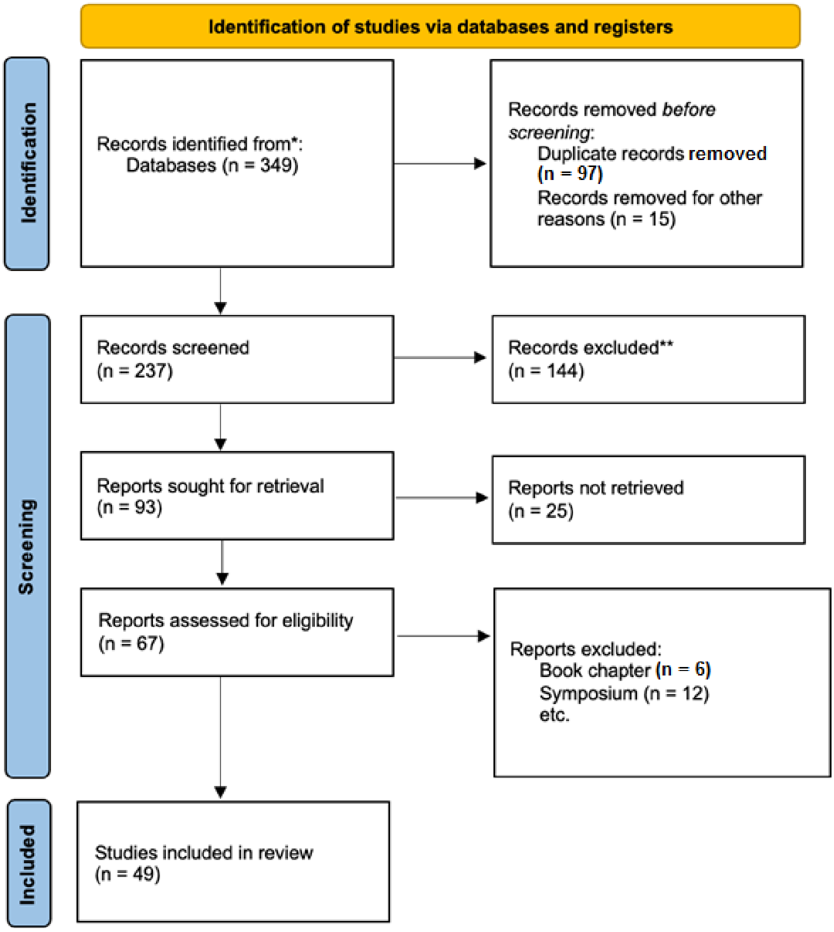

Figure 1. PRISMA Follow Diagram.

ers. The title of the questionnaire was used as the keyword. Besides that, articles that cited the questionnaire development articles were investigated.

Search Strategy

The study selection process is shown in **Figure 1** (Moher et al., 2009). As can be seen in the figure, the first step is database scanning. Studies applying the computational thinking scale developed by Korkmaz (2017) were determined by searching Google Scholar Academic, ProQuest, EBSCOhost, ERIC, ScienceDirect, Web of Science, and Taylor and Francis online databases. Searches of these databases have been conducted since 2015 when the "computational thinking'' scale was developed and published. Besides that, the snowball technique was also used to reach the studies: the studies that citateted the article at their reference list were also investigated. According to Korkmaz et al. (2017) and Korkmaz et al. (2015), all referenced studies were identified and it was

Table 1. Descriptive Features of the Included Studies.

	Frequency	%
Year		
2017	1	1.9
2018	8	15.4
2019	7	13.5
2020	17	32.7
2021	19	36.5
Item Number		
15 item	2	3.8
22 item	22	40.4
27 item	1	1.9
29 item	30	53.8
Language		
English	32	61.5
Turkish	19	36.5
Country		
Canada	1	1.9
China	2	3.8
Greece	2	3.8
India	1	1.9
Jakarta	1	1.9
Korea	1	1.9
Malaysia	1	1.9
Spain	1	1.9
Taiwan	3	5.8
TR	35	65.4
USA	1	1.9
Gender of Participants		
Female	6,064	52.76
Male	5,430	42.24

seen that there were 349 studies in total. These studies were reviewed according to the research inclusion criteria.

The study group of the research seen in the Prisma flowchart consisted of 49 studies that comply with the criteria determined in this study by the researchers. In addition, for studies reporting more than one reliability coefficient for different study groups, each coefficient was coded separately. As a result, 52 Cronbach's alpha coefficients were obtained from 49 studies (see **Appendix A**). The descriptive features of the studies are presented in **Table 1**.

Inclusion and Exclusion Criteria

To include articles in this RG study, the following criteria were considered: (i) the study had to be published, (ii) the study should be written in English or in Turkish, and (iii) the study had to report a reliability coefficient. Some studies reported the reliability coefficient of the original study instead of calculating a new one with their sample. Hence, they were not included in this meta-analysis.

Data Analysis and Interpretation

Reliability and validity are two important concepts of measurement. Reliability is related to the accuracy of the measurement process and the consistency of the scores of the same sample from the same test in two applications (Anastasi, 1976; McDonald, 1999). Depending on the sample characteristics, it can take different values for different samples despite using the same measurement instrument (Özdemir et al., 2020). The source and amount of these differences in the reliability coefficients are determined by RG analysis (Graham et al., 2006; Vacha-Haase, 1998).

The Cronbach alpha coefficient (Cronbach, 1951) is one of the most widely used reliability coefficients in social sciences (Bollen, 1989; Bonett & Wright, 2015). However, the use of the Cronbach alpha metric in the meta-analysis violates the normality assumption (Rodriguez & Maeda, 2006). To calculate Cronbach's alpha mean coefficient, it is necessary to normalize the distribution and stabilize its variances (Yoon et al., 2021). In this study, Bonett's (2002) transformation formula, presented below, was used.

$$Li = Ln\,(1 - \alpha i)$$

Li = Bonett metric, αi = Cronbach alpha, Ln is the natural logarithm

After applying the Bonett transform, reliability synthesis was made and the obtained values were converted back to the original Cronbach's alpha coefficient metric.

The heterogeneity between the reliability values of the studies included in the study was evaluated by calculating the Cochrane Q statistics and I2 index. Q statistics were applied to test the homogeneity assumption between alpha coefficients. The I2 index is a measure of heterogeneity (Higgins & Thompson, 2002). To interpret the results, the average effect sizes obtained with the Bonett transform, the lower and upper classes of the confidence interval, and the transformed Cronbach's alpha metric were used. Additionally, they were evaluated according to the 0.70 criterion (Nunnally & Bernstein, 1994).

Significant heterogeneity was seen among the estimates of the reliability of the primary studies that were included in the RG of CT. Analogue ANOVA analysis with moderator analysis was performed to investigate the sources of this heterogeneity. These moderator variables are sample type, test length, country, and language.

No matter how appropriate the methodology, the bias of the studies included in the meta-analysis may reduce the validity of the results (Becker, 2005). To eliminate this risk, the funnel plot asymmetry test, Egger's regression test, Begg and Mazumdar's rank correlation test, and Rosenthal's error protection number method were used to determine the publication bias status of primary studies. Comprehensive Meta-Analysis V2 free trial version was used for the aforementioned analyses.

Results

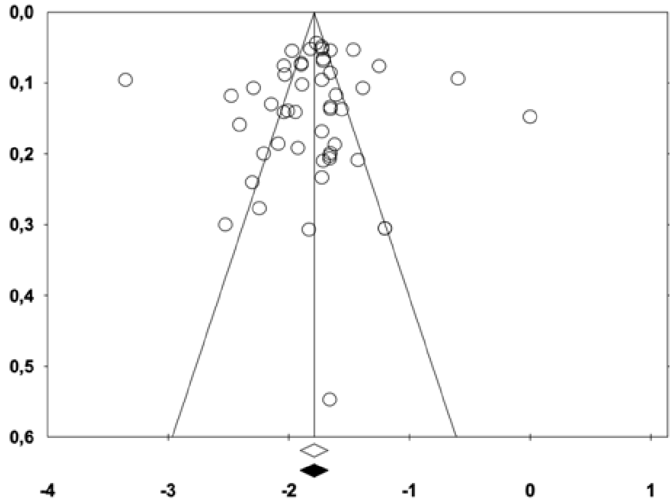

Figure 2. The Funnel Scatters Plot.

Table 2. Egger Regression Test results.

Regression line constant	SE	%95 CI		t	SD	P	
		UL	LL			(1-tailed)	(2-tailed)
-0.475	1.194	-2.88	1.93	0.398	48	0.346	0.692

UL: Upper limit; LL: Lower limit; SE: Standard error; SD: Standard deviation.

Publication Bias

In this study, the funnel scatter plot of the studies included in the meta-analysis, that is, the 52 observed studies, is exposed. The presence of tiny circles representing primary studies in the funnel means that primary studies are gathered around the average effect size. The fact that the studies are located symmetrically around the mean effect size in the funnel scatter plot shows that there is a meta-analysis without publication bias (**Figure 2**).

According to the Egger regression test, there is no bias in primary studies ($t(48) = 0.398$; $P = 0.692$, $P < 0.05$), but the finding needs to be tested with other methods as well (**Table 2**). If the Egger regression test and the funnel plot are evaluated together, the funnel plot of research without publication bias is symmetrical and the regression line is expected to coincide with the line in the center of the funnel plot (Sterne et al., 2005).

According to the Egger regression test applied to the reliability values of the primary studies included in the RG, the upper limit of the 95% confidence interval was

Table 3. Begg and Mazumdar Rank Correlation Test Results for Studies Used Korkmaz's (2017) CT Scale.

	Hypothesis
Kendall's S statistic (P-Q)	-55
Kendall's tau without continuity correction	
Tau	-0.04493
z-value for tau	0.46007
P-value (1-tailed)	0.32273
P-value (2-tailed)	0.64547
Kendall's tau with continuity correction	
Tau	-0.04412
z-value for tau	0.4517
P-value (1-tailed)	0.32574
P-value (2-tailed)	0.65148

Table 4. Rosenthal Error Protection Number Test Results for Studies Used Korkmaz's (2017) CT Scale.

Observed Studies						
Z	P	α	Z for α	Tails	# of Observed Studies	The number of studies that can ensure that P is greater than 0.05
-116.729	> 0.001	0.050	1.959	2.000	50	7,301

-2.88 and the limit inferior was 1.93. The value calculated for the regression line constant of 95 out of 100 studies to be conducted with other similar samples using the computational thinking scale developed by Korkmaz (2017) will be between the upper and lower limits of the 95% confidence interval.

It is also important to evaluate the Begg and Mazumdar rank correlation results, which quantify the result of the funnel plot, instead of the interpretation of the funnel plot, which depends on the researcher. The rank correlation between the standardized effect sizes and the standard errors of these values is 0, indicating that there is no publication bias (Hunter & Schmidt, 2004).

Begg and Mazumdar rank correlation tests were applied to statistically test that the study sample selected was not biased (**Table 3**). According to the Begg and Mazumdar rank correlation test, the publication bias test does not make sense ($\tau = -0.045$; $P = 0.65$). According to the analysis results shown in the chart, the primary studies included in this meta-analysis study were not biased. Rosenthal (1979) calculated the error-proof number (N) to find the summary of statistical significance (p-value) when there was little or no heterogeneity among the effect sizes of primary studies.

The Rosenthal fail-safe number calculated for primary investigations is 7,301 (**Table 4**). This value is statistically significant since $P < 0.05$. This figure states that for $P > 0.05$, 7,301 studies should be conducted in the opposite direction of the studies included in the meta-analysis with a zero-effect size value. A ratio $[1 < N / (5k+10)]$ be-

Table 5. Results on the Estimate of Pooled Reliability of the Computational Thinking Scale and Subscales.

	α	SE	Z	LL	UL	df	Q	I^2
Computational Thinking	0.843	0.065	-28.551*	0.823	0.863	49	1093.756*	95.520
Creativity	0.799	0.078	-20.618	0.765	0.827	32	822.182*	96.108
Algorithmic Thinking	0.848	0.064	-29.677*	0.828	0.866	32	532.432*	93.990
Cooperation	0.863	0.056	-33.956	0.833	0.866	32	420.606*	92.392
Critical Thinking	0.799	0.055	-28.920*	0.776	0.820	34	433.779*	92.162
Problem Solving	0.817	0.068	-25.104*	0.791	0.839	34	665.07*	94.888

*Note: k = number of reliability coefficients; SE = standard error; 95% confidence interval upper limit (UL) and lower limit (LL); df = degrees of freedom; Q = heterogeneity statistic representing total variance; I2 = heterogeneity index; *p < 0.05.*

tween Rosenthal's calculated error protection number (N) and the number of primary studies in the meta-analysis (k) has been proposed (Mullen, et al., 2001) at 28, indicating that the results of the meta-analysis are robust against publication bias and that the overall reliability estimate of primary studies is unlikely to affect the combined alpha coefficient of the current meta-analysis.

Data from the funnel plot, the Egger regression test, Beg and Mazumdar's rank correlation, and Rosenthal's fail-safe N tests show the absence of publication bias. Considering the results of all analyzes, it was determined that publication bias is not a serious threat affecting the RG estimation.

Results of the Estimate of the Pooled Reliability of the Computational Thinking Scale and Subscales

The Cronbach's alpha (α) coefficients and heterogeneity test results in the random-effects model are shown in **Table 1**. Cochrane Q parameter and Higgins I2 values were calculated to test the heterogeneity of primary studies. It was decided to use the random-effects model in the meta-analysis according to the results of heterogeneity tests.

As seen in **Table 5**, the Q (df = 49) value for the total CT scale was found to be 1093,756 (P < 0.001). The Q value calculated according to the chi-square (X^2) table is above the limit, with 49 degrees of freedom and a confidence interval of 0.05 (df = 49; X^2(0.05) = 66.339). Therefore, it can be said that there is heterogeneity among studies. The I^2 value (95.520) calculated for the heterogeneity test indicates a high level of heterogeneity. The random-effects model was used in this study as there was heterogeneity among primary studies. The overall Cronbach's alpha coefficient of the total CT scale was calculated as.843 (95%, 0.823 – 0.863) according to the random-effects model.

The meta-analysis of the RG of the CT scale is presented in the forest plot in **Figure 3**. When **Figure 3** is examined, it is seen that the reliability coefficients calculated in the Bonett type of all primary studies included in the analysis are significant. It was also observed that the Cronbach's alpha values of primary studies were between 0.70 and 0.965.

Heterogeneity tests were also performed for the sub-dimensions of the scale and creativity (Q (df = 32) = 822.182), algorithmic thinking (Q (df = 32) = 0.432), cooperation (Q (df = 32) = 420.606), critical thinking (Q (df = 34) = 433.779), and prob-

Study name	Fisher's Z	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Korkmaz_RG_22	-3,352	0,096	0,009	-3,540	-3,165	-35,017	0,000
Korkmaz_RG_44	-3,219	0,072	0,005	-3,359	-3,079	-44,953	0,000
Korkmaz_RG_18	-2,526	0,300	0,090	-3,114	-1,938	-8,416	0,000
Korkmaz_RG_23	-2,477	0,118	0,014	-2,709	-2,245	-20,958	0,000
Korkmaz_RG_21	-2,408	0,159	0,025	-2,719	-2,096	-15,150	0,000
Korkmaz_RG_24	-2,303	0,240	0,058	-2,773	-1,832	-9,586	0,000
Korkmaz_RG_20	-2,293	0,107	0,011	-2,502	-2,083	-21,426	0,000
Korkmaz_RG_31	-2,244	0,277	0,077	-2,787	-1,701	-8,103	0,000
Korkmaz_RG_45	-2,207	0,200	0,040	-2,598	-1,816	-11,059	0,000
Korkmaz_RG_14	-2,146	0,130	0,017	-2,400	-1,891	-16,533	0,000
Korkmaz_RG_15	-2,087	0,186	0,035	-2,452	-1,723	-11,235	0,000
Korkmaz_RG_11	-2,040	0,141	0,020	-2,317	-1,763	-14,443	0,000
Korkmaz_RG_32	-2,040	0,076	0,006	-2,188	-1,892	-27,008	0,000
Korkmaz_RG_49	-2,033	0,088	0,008	-2,205	-1,860	-23,076	0,000
Korkmaz_RG_39	-2,010	0,139	0,019	-2,283	-1,737	-14,430	0,000
Korkmaz_RG_43	-1,973	0,055	0,003	-2,080	-1,866	-36,094	0,000
Korkmaz_RG_9	-1,945	0,141	0,020	-2,222	-1,668	-13,781	0,000
Korkmaz_RG_2	-1,924	0,192	0,037	-2,300	-1,548	-10,036	0,000
Korkmaz_RG_7	-1,897	0,072	0,005	-2,038	-1,756	-26,311	0,000
Korkmaz_RG_25	-1,897	0,074	0,006	-2,043	-1,751	-25,526	0,000
Korkmaz_RG_51	-1,890	0,102	0,010	-2,091	-1,690	-18,516	0,000
Korkmaz_RG_41	-1,833	0,307	0,094	-2,434	-1,231	-5,972	0,000
Korkmaz_RG_40	-1,820	0,052	0,003	-1,922	-1,718	-34,961	0,000
Korkmaz_RG_34	-1,772	0,044	0,002	-1,858	-1,686	-40,527	0,000
Korkmaz_RG_6	-1,726	0,096	0,009	-1,914	-1,538	-18,028	0,000
Korkmaz_RG_17	-1,726	0,233	0,055	-2,184	-1,268	-7,392	0,000
Korkmaz_RG_19	-1,726	0,168	0,028	-2,056	-1,396	-10,246	0,000
Korkmaz_RG_27	-1,726	0,052	0,003	-1,828	-1,624	-33,169	0,000
Korkmaz_RG_50	-1,726	0,049	0,002	-1,822	-1,630	-35,331	0,000
Korkmaz_RG_16	-1,715	0,069	0,005	-1,850	-1,580	-24,935	0,000
Korkmaz_RG_26	-1,715	0,066	0,004	-1,844	-1,586	-26,076	0,000
Korkmaz_RG_35	-1,715	0,210	0,044	-2,126	-1,303	-8,168	0,000
Korkmaz_RG_5	-1,661	0,203	0,041	-2,058	-1,263	-8,193	0,000
Korkmaz_RG_10	-1,661	0,207	0,043	-2,066	-1,255	-8,031	0,000
Korkmaz_RG_33	-1,661	0,547	0,299	-2,733	-0,588	-3,036	0,002
Korkmaz_RG_3	-1,655	0,199	0,040	-2,045	-1,266	-8,326	0,000
Korkmaz_RG_4	-1,655	0,137	0,019	-1,924	-1,387	-12,104	0,000
Korkmaz_RG_29	-1,655	0,054	0,003	-1,761	-1,550	-30,688	0,000
Korkmaz_RG_37	-1,655	0,085	0,007	-1,823	-1,488	-19,375	0,000
Korkmaz_RG_42	-1,655	0,134	0,018	-1,918	-1,393	-12,371	0,000
Korkmaz_RG_1	-1,619	0,187	0,035	-1,986	-1,253	-8,666	0,000
Korkmaz_RG_36	-1,609	0,117	0,014	-1,839	-1,380	-13,741	0,000
Korkmaz_RG_48	-1,561	0,137	0,019	-1,830	-1,292	-11,373	0,000
Korkmaz_RG_28	-1,465	0,053	0,003	-1,570	-1,361	-27,508	0,000
Korkmaz_RG_13	-1,427	0,209	0,044	-1,837	-1,018	-6,831	0,000
Korkmaz_RG_30	-1,386	0,107	0,011	-1,596	-1,177	-12,959	0,000
Korkmaz_RG_12	-1,252	0,076	0,006	-1,401	-1,102	-16,431	0,000
Korkmaz_RG_46	-1,204	0,305	0,093	-1,802	-0,606	-3,944	0,000
Korkmaz_RG_47	-1,204	0,305	0,093	-1,802	-0,606	-3,944	0,000
Korkmaz_RG_8	-0,598	0,094	0,009	-0,782	-0,414	-6,368	0,000
	-1,831	0,013	0,000	-1,856	-1,805	-140,090	0,000

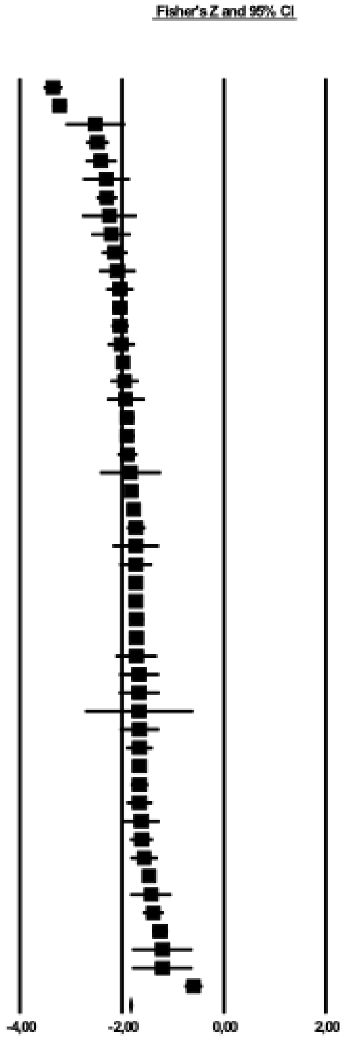

Figure 3. Forest Plot for CT Scale.

lem-solving ((Q (df = 34) = 665.07). 92.392; $I^2_{\text{Critical Thinking}}$ = 93.162%; $I^2_{\text{Problem Solving}}$ = 94.888%) indicate a high level of heterogeneity. As a result of the heterogeneity tests of the sub-dimensions, it was decided to use the random effects model for each dimension. According to the Cronbach's alpha coefficient random-effects model, creativity was calculated as 0.799 (95%, 0.765 – 0.827), algorithmic thinking 0.848 (95%, 0.828 – 0.866), cooperation 0.863 (95%, 0.833 – 0.0866), critical thinking 0.799 (95%, 0.776 – 0.820), and problem–solving was calculated as 0.817 (95%, 0.791 – 0.839).

The Relationship between Moderating Variables and Reliability Estimates

Table 6. The effect of Moderator Variables on the Reliability Estimate.

Variable	Category	k (%)	α	95%		Q_B	df	P
Test Length	15	2 (4)	0.700	0.370	0.857	4.667	3	0.198
	22	21 (40)	0.835	0.798	0.865			
	27	1 (2)	0.900	0.733	0.963			
	29	26 (54)	0.854	0.826	0.878			
Sample Type	Middle school	18 (36)	0.844	0.803	0.877	0.081	3	0.994
	High school	6 (16)	0.849	0.777	0.898			
	Undergraduate	22 (46)	0.842	0.805	0.872			
	Teachers	2 (2)	0.854	0.715	0.925			
Country	Asia	2 (4)	0.785	0.621	0.878	1.751	4	0.781
	Europe	3 (6)	0.800	0.652	0.885			
	Turkey	35 (73)	0.838	0.814	0.860			
	Far East	8 (17)	0.823	0.760	0.869			
Language	English	30 (63)	0.857	0.832	0.879	2.909	1	0.088
	Turkish	18 (37)	0.819	0.775	0.854			

The primary investigations of the study appear to show significant heterogeneity. It is important to identify the source of this heterogeneity. Sample type, language (categorical), test length, and (continuous) moderator of primary studies are examined as variables. Differences between subgroups are analyzed using the analogous ANOVA approach. To examine whether subgroups of all moderator variables were equally reliable, they were analyzed according to the mixed-effects model.

As seen in **Table 6**, it was determined that sample type, test length, country, and language variables did not affect the reliability of the measurements; in other words, there was no statistically significant difference between the subgroups ($P_{Test\ Length}$ = 0.198; $P_{Sample\ type}$ = 0.994; $P_{Country}$ = 0.781; $P_{Language}$ = 0.088).

Discussion

The present study aimed to investigate the overall reliability of the CT scale and the variability that can be explained by the moderators via the RG of meta-analysis. The CT scale was developed by Korkmaz et al. (2017) and adapted for different ages and languages (e.g. Korkmaz & Bai, 2019; Korkmaz & Özden, 2015). There are five subdimensions of CT in the scale: creativity, algorithmic thinking, cooperation, critical thinking, and problem-solving. To achieve these aims, research studies that used a CT scale and reported the reliability coefficient were examined. The reliability estimates of primary research showed high variability (I^2 = 95.52). This implies that the reliability of CT scale scores cannot be generalized to different contexts or populations. Furthermore, the variability, which was determined as I2, refers to the real differences among studies. Hence, the RG would be useful for the questionnaire (Graham et al., 2006; Vacha-Haase, 1998). Moreover, since I2 is significant, the random effect method was used while calculating the reliability coefficient across the reliability estimations (Borenstein et al., 2009).

Fifty-five studies used the CT scale and reported the reliability coefficient of the scores. The average reliability coefficient was calculated as 0.840. Besides that, a reliability estimation was also calculated for each dimension and found 0.799 for creativity, 0.848 for algorithmic thinking, 0.863 for cooperation, 0.799 for critical thinking, and.817 for problem-solving. Hence, the findings of the current study suggest that the reliability coefficients not only for the overall scale but also for sub-dimensions were within the acceptable range (Clark & Watson, 1995; DeVellis, 1991; Nunnally & Bernstein, 1994).

To investigate the variability of scores from the subcategories, the moderator analysis was conducted by using mixed-effect models. There were four different variables as moderators: the language of the article (English versus Turkish), test length (15, 22, 27 versus 29 items), country of study (Asia, Europe, Turkey, or Far East countries), and sample type (middle-school, high school, undergraduate, and teachers). As seen in **Table 6**, none of the moderator variables was significant, which means that there were no significant differences among subgroups in terms of the reliability coefficient. Regarding the effect of the sample type, previous research also suggested no reliability differences among groups. For instance, Hess et al. (2014) compared the reliability coefficients of perceived ease of use, perceived usefulness, and behavioral intentions of questionnaires for students and non-students and the findings of the study supported that there was no moderation effect of the sample type. Concerning the length of the scale, the findings of previous research are mixed. While some of them offer a significant effect (e.g. Streiner, 2003), others suggest no significant difference between the short or long form of the scale (e.g. Negrin et al., 2017). Supporting this discrepancy Niemi et al. (1986) investigated the impact of the length of the scale on reliability and recommended that reliability does not always increase with the number of items. Last but not least, some studies also propose no significant effect on either the language of the scale or the country where data was collected (e.g. Sen, 2021).

There are some limitations of the current study, including the conversion of Cronbach's alpha coefficients to Bonett's (2002) parameter, performing the analysis in the CMA program, and examining the sample type, test length, country, and language variables as sources of reliability measurement errors. Different reliability parameters in the literature consider different sources of measurement error. In future research, reliability estimates using other reliability parameters such as Hakstian and Whalen (1976) and Fisher Z can be evaluated by examining them. Furthermore, depending on whether it belongs to the definition and measurement, if RG has not been done, the reliability value of each study must be reported. However, in the database scanning process of this meta-analysis, many studies included the reliability values reported by Korkmaz (2017) instead of reporting the reliability of their measurements. This has reduced the number of primary studies that can be included in the RG study. Besides that, it was seen that the applications of the primary research of the study were mostly made in computer science and science teaching-related fields. CT is explained as the thought process involved in solving a problem (Curzon & McOwan, 2018; Denning, 2017). Considering that these intellectual processes can also be encountered in the solution of problems in different fields, it is necessary to consider CT in the teaching of different fields.

To sum up, this study supported the RG of the CT scale. At this point, it may be useful to remember that reliability is defined as the level of explanation of the phenomenon related to the conceptual structure that is expected to be measured with a scale (Şencan, 2005). However, reliability is a value of the data, not the scale (APA, 2020; Vacha-Haase, 1998). The reliability and validity of the scores obtained from the measurement tools must be investigated (Smith, 1987). Researchers are recommended to refer to a published RG study of the scale used or to report the reliability values of their measurements.

References

Aho, A. V. (2012). Computation and computational thinking. *The Computer Journal*, 55(7):833-835. DOI: https://doi.org/10.1093/comjnl/bxs074

Akman, E., & Bircan, M. A. (2021). Öğrencilerin Teknolojiyle Kendi Kendine Öğrenme ve Bilgisayarca Düşünme Becerilerinin İncelenmes i. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 40(1):12-22. DOI: https://doi.org/10.7822/omuefd.860638

Anastasi, A. (1976). Psychological Testing. In Group Analysis (Fourth Edi). Macmillan Publishing. DOI: https://doi.org/10.1177/053331647600900215

APA. (2020). Publication manual of the american psychological association. in psychological review (7th ed., Vol. 126, Issue 1). The Official Guide to APA Style. DOI: https://doi.org/10.1037/rev0000126

Avcu, Y. E., & Ayverdi, L. (2020). Examination of the Computer Programming Self-Efficacy's Prediction towards the Computational Thinking Skills of the Gifted and Talented Students. *International Journal of Educational Methodology*, 6(2):259-270. DOI: https://doi.org/10.12973/ijem.6.2.259

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, 27(4):335-340. DOI: https://doi.org/10.3102/10769986027004335

Çakır, E., & Yaman, S. (2018). Ters Yüz Sınıf Modelinin Öğrencilerin Fen Başarısı ve Bilgisayarca Düşünme Becerileri Üzerine Etkisi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 38(1):75-99.

Çevik, M., Barış, N., Şirin, M., Ortak, Ö. K., Kaplan, Y., Özdemir, B. A., Yalçın, H., Şeref, G., Topal, S., & Delice, T. (2021). The effect of digital activities on the technology awareness and computational thinking skills of gifted students (eTwinning project example). In International Journal of Modern Education Studies (Vol. 5, Issue 1). Available at: http://www.ijonmes.net

Çiftçi, S., Cengel, M., & Paf, M. (2018). Reflective thinking skills on computational thinking and problem solving as a predictor of self-efficacy of informatics teacher candidates on programming. *Kirşehir Eğitim Fakültesi Dergisi*, 19(1):335-348. DOI: https://doi.org/10.29299/kefad.2018.19.009

Clark, D. I. (2015). Computational and algorithmic thinking 2011-2015. AMT Publishing.

Crocker, L., & Algina, J. (2008). Introduction to classical and modern test theory- Procedures for Estimating Reliability. Cengage Learning.

Curzon, P., & McOwan, P. W. (2018). Computational thinking. In Springer. Available at: http://www.earticle.net/article.aspx?sn=249447

Denning, P. J. (2017). Computational thinking in science. *American Scientist*, 205(1):67-77. DOI: https://doi.org/10.1511/2017.124.13

Eser, M. T., & Aksu, G. (2021). Beck depression inventory-II: A study for Meta analytical reliability generalization. *Pegem Egitim ve Ogretim Dergisi*, 11(3):88-101. DOI: https://doi.org/10.14527/pegegog.2021.00

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1):38-43. DOI: https://doi.org/10.3102/0013189X12463051

Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, 41(2):219-231. DOI: https://doi.org/10.1007/BF02291840

Hess, T. J., McNab, A. L., & Basoglu, K. A. (2014). Reliability generalization of perceived ease of use, perceived usefulness, and behavioral intentions. *MIS Quarterly*, 38(1):1-28. DOI: https://www.jstor.org/stable/26554866

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414):557-560. DOI: https://doi.org/10.1136/bmj.327.7414.557

International Society for Technology in Education [ISTE] (2019). International society for technology in education standards. Available at: https://www.iste.org

Karaçaltı, C., & Korkmaz, Ö. (2018). Öğrencilerin Programlama Başarılarının Bilgisayarca-Eleştirel Düşünme ile Problem Çözme Becerileri Çerçevesinde İncelenmesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 7(2):343-370. DOI: https://dergipark.org.tr/tr/pub/amauefd/issue/41157/413487

Kıyıcı, G., & Yamak, H. (2021). Meta-analysis of computational thinking studies in science education: Descriptive statistics. *Sakarya University Journal of Education*, 11(3):526-544. DOI: https://doi.org/10.19126/suje.975108

Korkmaz, Ö., & Bai, X. (2019). Adapting computational thinking scale (CTS) for chinese high school students and their thinking scale skills level. *Participatory Educational Research*, 6(1):10-26. DOI: https://doi.org/10.17275/per.19.2.6.1

Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior*, 72:558-569. DOI: https://doi.org/10.1016/j.chb.2017.01.005

Korkmaz, Ozden, & Özden, M. Y. (2015). Bilgisayarca Düşünme Beceri Düzeyleri Ölçeğinin (BDBD) Ortaokul Düzeyine Uyarlanması. *Gazi Eğitim Bilimleri Dergisi*, 1:67-86. DOI: https://www.researchgate.net/publication/291338987

Korkmaz, Ozgen, Cakır, R., & Ozden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior*, 72:558-569. DOI: https://doi.org/10.1016/j.chb.2017.01.005

Lei, H., Ming, M., Li, F., Wang, X., & Geng, Y. (2020). Children and Youth Services Review Computational thinking and academic achievement: A meta-analysis among students. Children and Youth Services Review, 118:105439. DOI: https://doi.org/10.1016/j.childyouth.2020.105439

Liang, J., Shou, Y., Wang, M. C., Deng, J., & Luo, J. (2021). Alabama Parenting Questionnaire-9: A reliability generalization meta-analysis. *Psychological Assessment*, 33(10):940-951. DOI: https://doi.org/10.1037/pas0001031

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, 6(7):e1000100. DOI: https://doi.org/10.1371/journal.pmed.1000100

McDonald, R. P. (1999). Test theory: A unified treatment. *Journal of the American Statistical Association*, 95(451):1012-1013. DOI: https://doi.org/10.2307/2669496

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7):e1000097. DOI: https://doi.org/10.1371/journal.pmed.1000097

Negrín, F., Hernández-Fernaud, E., Hess, S., & Hernández, B. (2017). Discrimination of urban spaces with different level of restorativeness based on the original and on a shorter version of hartig et al.'s perceived restorativeness scale. *Frontiers in psychology*, 8:1735. DOI: https://doi.org/10.3389/fpsyg.2017.01735

Niemi, R. G., Carmines, E. G., & McIver, J. P. (1986). The impact of scale length on reliability and validity. *Quality and Quantity*, 20(4):371-376. DOI: https://doi.org/10.1007/BF00123086

Olderbak, S., Riggenmann, O., Wilhelm, O., & Doebler, P. (2021). Reliability generalization of tasks and recommendations for assessing the ability to perceive facial expressions of emotion. *Psychological Assessment*, 33(10):911-926. DOI:

https://doi.org/10.1037/pas0001030
Oluk, A., Korkmaz, Ö., & Oluk, H. A. (2018). Effect of scratch on 5th graders' algorithm development and computational thinking skills. *Turkish Journal of Computer and Mathematics Education*, 9(1):54-71. DOI: https://doi.org/10.16949/turkbilmat.399588
Opitz, M. C., Newman, E., Alvarado Vázquez Mellado, A. S., Robertson, M. D. A., & Sharpe, H. (2020). The psychometric properties of Orthorexia Nervosa assessment scales: A systematic review and reliability generalization. *Appetite*, 155:104797. DOI: https://doi.org/10.1016/j.appet.2020.104797
Özgür, H. (2020). Relationships between computational thinking skills, ways of thinking and demographic variables: A structural equation modeling. *International Journal of Research in Education and Science*, 6(2):299-314. DOI: https://doi.org/10.46328/ijres.v6i2.862
Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A.C., Welch, V. A., Whiting, P., Mckenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *The BMJ*, 372:n160. DOI: https://doi.org/10.1136/bmj.n160
Papert, S. (1996). An exploration in the sapace of mathematics educations. *International Journal of Computers for Mathematical Learning*, 1(1):95-123. DOI: https://doi.org/10.1007/bf00191473
Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior*, 72:678-691. DOI: https://doi.org/10.1016/j.chb.2016.08.047
Sen, S. (2021). A reliability generalization Meta-analysis of Runco ideational behavior scale. *Creativity Research Journal*, 33:1-17. DOI: https://doi.org/10.1080/10400419.2021.1960719
Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1):99-103.
Şencan, H. (2005). Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik (1st ed.).
Tsai, M. J., Liang, J. C., & Hsu, C. Y. (2021). The computational thinking scale for computer literacy education. *Journal of Educational Computing Research*, 59(4):579-602. DOI: https://doi.org/10.1177/0735633120972356
Vacha-Haase, T. (1998). Reliability generalization: exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1):6. DOI: https://doi.org/10.1177/0013164498058001002
Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, 62(4):562-569. DOI: https://doi.org/10.1177/0013164402062004002
Vacha-Haase, T., Kogan, L. R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement*, 60(4):509-522. DOI: https://doi.org/10.1177/00131640021970682
Vaidyanathan, S. (2016). What's the Difference Between Coding and Computational Thinking? EdSurge. Available at: https://www.edsurge.com/news/2016-08-06-what-s-the-difference-between-coding-and-computational-thinking
Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8):594-604. DOI: https://doi.org/10.1037//0003-066x.54.8.594
Wing, J. M. (2006). Computational Thinking. *Communications of The Acm*, 49(3):33-35. DOI: https://doi.org/10.14742/ajet.3521
Yalçın, N., & İkinci, A. (2020). Meslek Liseleri Bilişim Teknolojileri Alan Öğrencilerinin Bilgisayarca Düşnme Beceri Düzeylerinin Eğitim Program Türüne Göre İncelenmesi.

Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 20(4):1639-1656. DOI: https://doi.org/2020.20.58249-650813

Yoon, S., Yang, Y., Ro, E., Ahn, W.-Y., Kim, J., Shin, S.-H., Chey, J., & Choi, K.-H. (2021). Reliability, and convergent and discriminant validity of gaming disorder scales: A meta-analysis. *Frontiers in Psychology*, 12:1-15. DOI: https://doi.org/10.3389/fpsyg.2021.764209

Received: 04 June 2022
Revised: 19 August 2022
Accepted: 30 August 2022

Appendix I

Studies Used in the Meta-Analysis

Akgün, F. (2020). Öğretmen Adaylarının Bilgi Ve İletişim Teknolojileri Yeterlikleri Ve Bilgi İşlemsel Düşünme Becerilerinin Çeşitli Değişkenler Açısından Değerlendirilmesi [An Evaluatıon On Pre-Servıce Teacher's Informatıon And Communıcatıons Technology Competency And Comp. Trakya University Journal of Social Science, 22(1), 629–654. https://doi.org/10.26468/trakyasobed.679581

Akman, E., & Bircan, M. A. (2021). Öğrencilerin Teknolojiyle Kendi Kendine Öğrenme ve Bilgisayarca Düşünme Becerilerinin İncelenmesi [Examination of Students' Self-directed Learning with Technology and Computational Thinking Skills]. Ondokuz Mayis University Journal of Education Faculty, 40(1), 12–22. https://doi.org/10.7822/omuefd.860638

Ata, R., & Çevik, M. (2020). Understanding predictor effects of computational thinking skills and media and technology use and attitudes of pre-service teachers for STEM awareness. KEDI Journal of Educational Policy, 17(1), 99–121.

Ata, R., & Yıldırım, K. (2020). Analysis of the relation between computational thinking and new media literacy skills of first-year engineering students. Journal of Educational Multimedia and Hypermedia, 29(1), 5–20. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc17&NEWS=N&AN=2020-06219-001

Atman Uslu, N., Mumcu, F., & EĞİN, F. (2018). Görsel Programlama Etkinliklerinin Ortaokul Öğrencilerinin Bilgi-İşlemsel Düşünme Becerilerine Etkisi [The Effect of Visual Programming Activities on Secondary School Students' Computational Thinking Skills]. Journal of Ege Education Technologies, 2(1), 19–31.

Avcu, Y. E., & Ayverdi, L. (2020). Examination of the Computer Programming Self-Efficacy's Prediction towards the Computational Thinking Skills of the Gifted and Talented Students. International Journal of Educational Methodology, 6(2), 259–270. https://doi.org/10.12973/ijem.6.2.259

Aydoğdu, Ş. (2020). Blok Tabanlı Programlama Etkinliklerinin Öğretmen Adaylarının Programlamaya İlişkin Öz Yeterlilik Algılarına Ve Hesaplamalı Düşünme Becerilerine Etkisi [The Effect Of Block-Based Programmıng Actıvıtıes On Pre-Servıce Teachers' Computer Programmıng Self-Ef. Educational Technology Theory And Practice, 10(1), 303–320. https://doi.org/https://doi.org/10.17943/etku.649585

Bilbao, J., Bravo, E., García, O., Rebollar, C., & Varela, C. (2021). Study to find out the perception that first year students in engineering have about the Computational Thinking skills, and to identify possible factors related to the ability of Abstraction. Heliyon, 7(2). https://doi.org/10.1016/j.heliyon.2021.e06135

Bulut, A. E., & Yılmaz, M. (2021). Fen Lisesi Öğrencilerinin Bilgi İşlemsel Düşünme Beceri Düzeylerinin Belirlenmesi [Determining the Computational Thinking Skill Levels of Science High School Students]. Gazi Journal of Education Sciences (GJES), 7(1), 80–91. https://doi.org/https://dx.doi.org/110.30855/gjes.2021.07.01.005

Cakır, E., & Yaman, S. (2018). Ters Yüz Sınıf Modelinin Öğrencilerin Fen Başarısı ve Bilgisayarca Düşünme Becerileri Üzerine Etkisi [The Effect of Flipped Classroom Model on Students' Science Success and Computational Thinking Skills]. Gazi University Journal of Gazi Education Faculty, 38(1), 75–99.

Çakır, R. (2021). Computational Thinking Skills of Turkish and Indian Teacher Candidates: A Comparative Study. International Journal of Psychology and Educational Studies, 8(1), 24–37. https://doi.org/10.17220/ijpes.2021.8.1.226

Çevik, M., Barış, N., Şirin, M., Ortak, Ö. K., Kaplan, Y., Özdemir, B. A., Yalçın, H., Şeref, G., Topal, S., & Delice, T. (2021). The effect of digital activities on the technology awareness and computational thinking skills of gifted students (eTwinning project example). In International Journal of Modern Education Studies (Vol. 5, Issue 1). http://www.ijonmes.net

Chen, Y. C., Tsui, P. L., & Lee, C. S. (2021). Is mathematics required for cooking? An interdisciplinary approach to integrating computational thinking in a culinary and restaurant management course. Mathematics, 9(18). https://doi.org/10.3390/math9182219

Doleck, T., Bazelais, P., Lemay, D. J., Saxena, A., & Basnet, R. B. (2017). Algorithmic thinking, cooperativity, creativity, critical thinking, and problem solving: exploring the relationship between computational thinking skills and academic performance. Journal of Computers in Education, 4(4), 355–369. https://doi.org/10.1007/s40692-017-0090-9

Durak, Hatice Yildiz, & Saritepeci, M. (2018). Analysis of the relation between computational thinking skills and various variables with the structural equation model. Computers and Education, 116, 191–202. https://doi.org/10.1016/j.compedu.2017.09.004

Durak, Hatice Yıldız. (2020). The Effects of Using Different Tools in Programming Teaching of Secondary School Students on Engagement, Computational Thinking and Reflective Thinking Skills for Problem Solving. Technology, Knowledge and Learning, 25, 179–195. https://doi.org/https://doi.org/10.1007/s10758-018-9391-y

Durak, Hatice Yıldız, Saritepeci, M., & Durak, A. (2021). Modeling of Relationship of Personal and Affective Variables With Computational Thinking and Programming. Technology, Knowledge and Learning, 0123456789. https://doi.org/10.1007/s10758-021-09565-8

En, L. K., Karpudewan, M., & Zaharudin, R. (2021). Computational thinking in STEM education among matriculation science students. Asia Pacific Journal of Educators and Education, 36(1), 177–194. https://doi.org/10.21315/apjee2021.36.1.10

Gong, D., Yang, H. H., & Cai, J. (2020). Exploring the key influencing factors on college students' computational thinking skills through flipped-classroom instruction. International Journal of Educational Technology in Higher Education, 17(1). https://doi.org/10.1186/s41239-020-00196-0

Günbatar, M. S. (2019). Computational thinking within the context of professional life: Change in CT skill from the viewpoint of teachers. Education and Information Technologies, 24(5), 2629–2652. https://doi.org/10.1007/s10639-019-09919-x

Günbatar, M. S., & Bakırcı, H. (2019). STEM teaching intention and computational thinking skills of pre-service teachers. Education and Information Technologies, 24(2), 1615–1629. https://doi.org/10.1007/s10639-018-9849-5

Hamutoğlu, N. B. (2020). The Effect of Information Technologies Lesson Designed in Constructivist Approach on the Computational Thinking and Problem Solving Skills of 6th Grader Refugee Students. Journal of Ahi Evran University Kırşehir Education Faculty, 21(2), 834–874. https://doi.org/10.29299/kefad.XXXX

Hava, K., & Ünlü, Z. (2021). Investigation of the Relationship Between Middle School Students' Computational Thinking Skills and their STEM Career Interest and Attitudes Toward Inquiry. Journal of Science Education and Technology, 484–495. https://doi.org/10.1007/s10956-020-09892-y

He, Z., Wu, X., Wang, Q., & Huang, C. (2021). Developing eighth-grade students' computational thinking with critical reflection. Sustainability (Switzerland), 13(20). https://doi.org/10.3390/su132011192

Hıdıroğlu, Y. Ö., & Hıdıroğlu, Ç. N. (2020). Matematik Öğretmenlerinin Zihin Tipleri ile Bilgisayarca Düşünme Becerileri Arasındaki İlişki [The Relationship between Mathematics Teachers' Mind Types and Computational Thinking Skills]. In Pamukkale University Journal of Education. https://doi.org/10.9779/pauefd.696511

Karaçaltı, C., & Korkmaz, Ö. (2018). Öğrencilerin Programlama Başarılarının Bilgisayarca-Eleştirel Düşünme ile Problem Çözme Becerileri Çerçevesinde İncelenmesi [Examination of the Students' Computational-

Critical Thinking and Problem-Solving Skills on their Success of Programming Course]. Amasya Education Journal, 7(2), 343–370. https://dergipark.org.tr/tr/pub/amauefd/issue/41157/413487

Kirmit, Ş., Dönmez, İ., & Çataltaş, H. E. (2018). Üstün Yetenkli Öğrencilerin Bilgisayarca Düşünme Becerilerinin İncelenmesi [The Study Of Gıfted Students' Computatıonal Thınkıng Skılls]. Journal Of STEAM Education, 2(1), 17–26. https://dergipark.org.tr/tr/pub/steam/issue/42077/476833

Koçak, Ö., & Göksu, İ. (2020). Examining 21st Century Skill Levels of Students and the Relationship between Skills. İnönü University Journal of the Faculty of Education, 21(20 20), 772–784. https://doi.org/10.17679/inuefd.656784

Korkmaz, Ö., Kaya, M., & Çakır, R. (2020). Oyunlaştırılmış Robot Etkinliklerinin Ortaokul Öğrencilerinin Problem Çözme ve Bilgi İşlemsel Düşünme Becerilerine Etkisi [The Effect of Gamified Robotics Activities on the Problem Solving and the Computational Thinking Skills of the Secondary School Stud. Journal of Ege Education, 21(1), 54–70. https://doi.org/10.12984/egeefd.588512

Lai, C. F., Zhong, H. X., & Chiu, P. S. (2021). Investigating the impact of a flipped programming course using the DT-CDIO approach. Computers and Education, 173(November 2020). https://doi.org/10.1016/j.compedu.2021.104287

Lee, M., & Lee, J. (2021). Enhancing computational thinking skills in informatics in secondary education: the case of South Korea. Educational Technology Research and Development, 69(5), 2869–2893. https://doi.org/10.1007/s11423-021-10035-2

Lin, P. H., & Chen, S. Y. (2020). Design and Evaluation of a Deep Learning Recommendation Based Augmented Reality System for Teaching Programming and Computational Thinking. IEEE Access, 8, 45689–45699. https://doi.org/10.1109/ACCESS.2020.2977679

Mindetbay, Y., Bokhove, C., & Woollard, J. (2019). What is the Relationship between Students' Computational Thinking Performance and School Achievement? International Journal of Computer Science Education in Schools, 2(5), 3–19. https://doi.org/10.21585/ijcses.v0i0.45

Muliyati, D., Sumardani, D., Siswoyo, S., Bakri, F., Permana, H., Handoko, E., & Sari, N. L. K. (2021). Development and evaluation of granular simulation for integrating computational thinking into computational physics courses. Education and Information Technologies, 27(2), 2585–2612. https://doi.org/10.1007/s10639-021-10724-8

Oluk, A., & Çakır, R. (2019). Investigating University Students ' Computational Thinking Skills in Terms of Logical Mathematical Intelligence Problem

Solving Skills. Journal of Theoretical Educational Science, 12(2), 457–473. https://doi.org/http://dx.doi.org/10.30831/akukeg.351312

Özgür, H. (2020). Relationships between computational thinking skills, ways of thinking and demographic variables: A structural equation modeling. International Journal of Research in Education and Science, 6(2), 299–314. https://doi.org/10.46328/ijres.v6i2.862

Paf, M., & Dinçer, B. (2021). A Study of the Relationship between Secondary School Students' Computational Thinking Skills and Creative Problem-Solving Skills. Turkish Online Journal of Educational Technology, 20(4), 1–15. https://www.researchgate.net/publication/355219924_A_Study_of_the_Relationship_between_Secondary_School_Students%27_Computational_Thinking_Skills_and_Creative_Problem-Solving_Skills

Pellas, N., & Vosinakis, S. (2018). The effect of simulation games on learning computer programming: A comparative study on high school students' learning performance by assessing computational problem-solving strategies. Education and Information Technologies, 23(6), 2423–2452. https://doi.org/10.1007/s10639-018-9724-4

Saritepeci, M. (2020). Developing Computational Thinking Skills of High School Students: Design-Based Learning Activities and Programming Tasks. Asia-Pacific Education Researcher, 29(1), 35–54. https://doi.org/10.1007/s40299-019-00480-2

Sarıtepeci, M. (2021). Modelling the Effect of TPACK and Computational Thinking on Classroom Management in Technology Enriched Courses. Technology, Knowledge and Learning, 0123456789. https://doi.org/10.1007/s10758-021-09529-y

Sirakaya, D. A. (2020). Investigating computational thinking skills based on different variables and determining the predictor variables. Participatory Educational Research, 7(2), 102–114. https://doi.org/10.17275/per.20.22.7.2

Sırakaya, D. A. (2020). Investigation of Computational Thinking in the Context of ICT and Mobile Technologies. International Journal of Computer Science Education in Schools, 3(4), 50–59. https://doi.org/10.21585/ijcses.v3i4.73

Sırakaya, M., Sırakaya, D. A., & Korkmaz, Ö. (2020). The Impact of STEM Attitude and Thinking Style on Computational Thinking Determined via Structural Equation Modeling. Journal of Science Education and Technology, 29(4), 561–572. https://doi.org/10.1007/s10956-020-09836-6

Tonbuloğlu, B., & Tonbuloğlu, I. (2019). The effect of unplugged coding activities on computational thinking skills of middle school students.

Informatics in Education, 18(2), 403–426. https://doi.org/10.15388/infedu.2019.19

Türker, P. M., & Pala, F. K. (2020). The Effect of Algorithm Education on Students' Computer Programming Self-Efficacy Perceptions and Computational Thinking Skills. International Journal of Computer Science Education in Schools, 3(3), 19–32. https://doi.org/10.21585/ijcses.v3i3.69

Yalçın, N., & İkinci, A. (2020). Meslek Liseleri Bilişim Teknolojileri Alan Öğrencilerinin Bilgisayarca Düşnme Beceri Düzeylerinin Eğitim Program Türüne Göre İncelenmesi. Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 20(4), 1639–1656. https://doi.org/2020.20.58249-650813

Zhang, J. H., Meng, B., Zou, L. C., Zhu, Y., & Hwang, G. J. (2021). Progressive flowchart development scaffolding to improve university students' computational thinking and programming self-efficacy. Interactive Learning Environments, 0(0), 1–18. https://doi.org/10.1080/10494820.2021.1943687