

The Effect of Goal-Based Scenario Curriculum on Computational Thinking Skills and Academic Achievement*

Veysel Karani Ceylan**, Ruken Akar Vural***

Makale Geliş Tarihi: 20/12/2024

Makale Kabul Tarihi: 06/05/2025

DOI: 10.35675/befdergi.1604842

Abstract

The research aims to examine the effect of the Scratch curriculum developed with Goal-Based Scenario Learning (GBSL) design on 6th-grade students' Computational Thinking (CT) skills and academic achievement according to different variables. The research model was a 2X3 quasi-experimental. The CT Self-Assessment Scale and Academic Achievement Test were used as data collection tools in the study. In data analysis, both parametric and parametric tests were used. The experimental group students performed better on academic achievement post-test and retention-test scores. The post-test and retention-test mean scores of the CT self-assessment scale did not differ between groups, but the abstraction dimension scores in the sub-dimension scores did differ in favor of the experimental group. While the post-test and retention-test mean scores of both group students in CT skills were similar, the experimental group's scores differed between test scores, and there was a significant difference between the groups only in abstraction sub-dimension scores.

Keywords: Academic achievement, computational thinking skill, goal-based scenario learning

Senaryo Temelli Öğretim Programının Bilgi İşlemsel Düşünme Becerisine ve Akademik Başarıya Etkisi

Öz

Araştırmada 6. sınıf bilişim teknolojiler dersi programlama ve yazılım geliştirme ünitesi Hedef Temelli Senaryo Öğrenme modeli kullanılarak yeniden geliştirilmiş ve bu öğretim programının öğrencilerin akademik başarılarına ve bilgi işlemsel düşünme becerilerine olan etkisi incelenmiştir. Araştırmada yöntem olarak nicel yaklaşımlardan 2X3 (deney ve kontrol grubu X ön test, son test ve kalıcılık testi) yarı-deneysel modeli tercih edilmiştir. Araştırmacı tarafından geliştirilen Bilgi İşlemsel Düşünme Öz Değerlendirme Ölçeği ve Akademik Başarı Testi veri toplama araçları olarak kullanılmıştır. Veri analizinde çok değişkenli kovaryans

* This study was produced from the doctoral thesis titled "The effect of scenario based learning Scratch curriculum on sstudents' computational thinking skills and learning outcomes of problem solving and programming unit" written by the first author under the supervision of the second author.

** Millî Eğitim Bakanlığı, Aydın, Türkiye, vceylan@gmail.com, Orcid: [0000-0002-8899-960X](https://orcid.org/0000-0002-8899-960X)



*** Aydın Adnan Menderes Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Aydın,

Türkiye, rakarvural@gmail.com, Orcid: [0000-0002-3137-3753](https://orcid.org/0000-0002-3137-3753)



Kaynak Gösterme: Ceylan, V. K. & Akar Vural, R. (2025). The effect of goal-based scenario curriculum on computational thinking skills and academic achievement. *Bayburt Eğitim Fakültesi Dergisi*, 20(47), 1038-1067.

analizi, parametrik olmayan Friedmann testi ve takip testleri kullanılmıştır. Araştırma bulgularında akademik başarı son test ve kalıcılık testi puanlarına göre deney grubu öğrencilerinin lehine anlamlı farklılık bulunmuştur. BID öz değerlendirme ölçeği son test ve kalıcılık testi ortalama puanlarına göre gruplar arasında anlamlı fark bulunmazken, alt boyut puanlarında gruplar arasında yalnızca soyutlama boyutu puanlarına göre deney grubu lehine anlamlı farklılığın olduğu tespit edilmiştir.

Anahtar Kelimeler: *Bilgi işlemsel düşünme, hedef temelli senaryo öğrenme, akademik başarı*

Introduction

Considering the area that artificial intelligence will occupy in the future, rather than just the tool that can be used at the end-user level, it can be easily predicted that individuals who can develop artificial intelligence will be needed. Among these skill sets, Computational Thinking (CT) has been prominently emphasized in recent years. While Papert first presented the concept of CT in 1980, it was computer scientist Wing who revitalized interest in the topic in 2006 by elaborating on its definition. Wing posits that CT is a sophisticated cognitive ability that enables humans to tackle problems, develop systems, and understand human behaviour via the framework of computer science principles. Moreover, Wing (2006) underscored that this skill is vital not only for those in the information technology field but for all persons. While there is no consensus regarding the fundamental skills of CT, an examination of various perspectives reveals that algorithmic thinking, abstraction, parallelization, and decomposition are widely recognized as agreed-upon sub-dimensions (Guzdial et al., 2019; ISTE, 2016; Csizmadia et al., 2015; Wing, 2006). CT skills are positively connected with reasoning, problem-solving, reflective thinking, and spatial thinking, which are classified as higher-order thinking domains (Yıldız-Durak & Sarıtepeci, 2018; Swanson, 2017; Selby & Woollard, 2014; Orton et al., 2005). CT includes abstraction, problem-solving, and algorithmic reasoning, which are acknowledged as vital higher-order cognitive skills (Berry, 2013; Brennan & Resnick, 2012).

As the significant impact of computer science on contemporary and future technology, it is essential to educate students about ICT, regarded as the foundational cognitive skill of this discipline, from the early years of education (Yadav et al., 2016). 2015 research by Google revealed that administrators, instructors, students, and parents in U.S. elementary and secondary schools believe that computer science curricula encompassing these abilities should be implemented. An efficient education in coding and programming is one of the most effective means for individuals to develop skills of CT (Hoppe & Manske, 2022; Joint Research Council [JRC], 2016). However, studies in the literature indicate that learners struggle with high-level cognitive concepts such as data structures, algorithms, and abstract thinking encountered in coding and programming education, leading to elevated dropout rates due to failures in these courses (Sivasakthi & Rajendran, 2011; Kelleher et al., 2007;

Kaasboll, 1998). The literature underscores that training with coding tools designed to visualize concepts and render them more tangible enhances success and engagement in programming education (Resnick et al., 2009; Arabacıoğlu et al., 2007; Gültekin, 2006).

Barr and Stephenson (2011) asserted that ambiguity persists in the instruction of CT, and there is an absence of a standardized teacher training program in this domain. Teachers of information technology who were in charge of putting together the Problem Solving and Programming unit said it was hard because the course was so new and there wasn't a single guide or resource that had all the information they needed about CT. In this setting, educators in the field persist in employing diverse individual applications rather than a holistic approach nationwide in the instruction of the subject (Kukul, 2018; Kalelioğlu et al., 2014).

The adoption of the constructivist approach in educational settings has led theorists to highlight the integration of thinking skills with content and the significance of educational contexts about meaning (Bransford et al., 1990; Brown et al., 1989). Schank (1994) developed a teaching framework known as Goal-Based Scenario Learning (GBSL) for this purpose. GBSL is a method that delineates a problem-solving role and task for learners to attain a clearly defined goal, providing a learning-by-doing environment (Foster & Bareiss, 1995). GBSL offers a strategy that connects students to the realities of their intended profession by constructing authentic learning experiences. In GBSL, learners are regarded as prospective professionals and are provided with a scenario descriptor or a realistic set of conditions. In this context, learners are supported in navigating dilemmas through one or more pre-formulated focus questions that exemplify various scenarios, aiding them in achieving the specified learning objectives. Learners frequently adopt specific roles or perspectives, allowing them to utilize scenarios in contexts that provide them with a competitive advantage. The scenario's resourcefulness and the teacher's strategic selection of focus questions enable students to demonstrate skills or processes, solve problems, comment on information, or explore topics (Errington, 2010). Numerous studies indicate that GBSL is an effective teaching method (Zumbach & Reimann, 2002; Schoenfeld-Tacher et al., 2001; Bell et al., 1993).

An examination of the literature reveals that while various tools and methods have been employed in coding and programming education (Top & Arabacıoğlu, 2024; Alp, 2019; Dinçer, 2018; Malan & Leitner, 2007), there is a scarcity of empirical studies integrating Goal-Based Scenario Learning Teaching with coding education and the enhancement of CT at the primary level (Bond & Bedenlier, 2019; Angeli et al., 2016; Arabacıoğlu, 2012).

This study aims to evaluate the effect of the scenario-based Scratch curriculum on students' academic achievement scores, retention in learning, and CT skills in the 6th-grade information technologies course problem-solving and programming unit. In line with this purpose, answers to the following research questions were sought;

1. Was there a significant difference (sig. dif.) in the total tests' scores of the CT Self-Assessment Scale (CTSAS) and its sub-dimensions (algorithmic thinking, parallelization, decomposition, abstraction, and automatization) when comparing the experimental (exp.) and control group students, favoring the exp. group both within and between the groups?
2. Was there a sig. dif. in post-test and retention-test scores favoring the exp. group after controlling for the pre-test results of the AAT between the exp. and control groups?
3. When controlling for the academic achievement pre-test scores of the exp. and control groups, is there a sig. dif. in the post-test and retention-test scores favoring the exp. group, considering gender, frequency of screen usage, and socioeconomic variables of the students?

Method

Experimental models are research frameworks in which the data observed are generated directly under the researcher's control to ascertain cause-and-effect correlations. (Karasar, 2013). Quasi-experimental models are models with high application validity that are used in many situations where the controls required by real experimental models cannot be provided. Since there is no possibility of experimental study in different groups in the MEB environment, a quasi-experimental model was used. In addition, in this model, the validity and reliability of the study were tried to be increased by applying pre-test, post-test and retention repeated tests to unequal groups.. The quasi-experimental model used in this study was designed according to the 2X3 (experiment, control group X pre-test, post-test, and retention-test) quasi-experimental model. In this frequently used model, there are pre-test and post-test applications in both groups. The factor that reduces the internal validity in the application of this design is the possibility that the differences detected as a result of the analyses are due to a pre-existing situation in the groups (Balçı, 1995:250). To increase the internal validity, it was tried to control the pre-existing differences by performing covariance analysis in data analysis.

Sampling

Stratified sampling is a sampling in which subgroups in the population are guaranteed to be represented in the sample (Balçı, 1995). A stratified sample was employed, incorporating sixth grades from three schools representing diverse socioeconomic strata in the Milas region of Muğla province during the 2018-2019 academic year. 122 students were selected from two 6th grades in each school, with one group as the control and the other as the exp. group. In the determination of the schools where the study will be carried out, schools with low, middle, and high socioeconomic status (two public and one private school), with an information technologies laboratory and an information technologies teacher, were determined. The sample of the research was shown in Table 1.

Table 1.
Sample of the Study

| | Experimental Group | | Control Group | | Total | |
|-----------------------------|--------------------|------|---------------|------|-------|-----|
| | n | % | n | % | n | % |
| High socioeconomic | 12 | 50 | 12 | 50 | 24 | 100 |
| Female | 5 | 41.7 | 4 | 33 | 9 | |
| Male | 7 | 58.3 | 8 | 67 | 15 | |
| Middle socioeconomic | 32 | 53.3 | 28 | 47.7 | 60 | 100 |
| Female | 16 | 50 | 9 | 32 | 25 | |
| Male | 16 | 50 | 19 | 68 | 35 | |
| Low socioeconomic | 19 | 50 | 19 | 50 | 38 | 100 |
| Female | 12 | 63 | 12 | 63 | 24 | |
| Male | 7 | 37 | 7 | 37 | 14 | |
| Total | 63 | 52 | 59 | 48 | 122 | 100 |

Upon analyzing Table 1, it was found that 66% of the students involved in the study were female, while 34% were male. The total number of experimental group students was 63, while the total number of control group students was 59. In determining the exp. and control groups in the schools, the fact that these classes had not received coding and programming education in the previous years, the level of their grade point averages in the last year, and the interviews with the teachers teaching their courses were taken into consideration.

Data Collection and Analysis

The study employed the CT Self-Assessment Scale (CTSAS), created by the researcher (Ceylan & Akar Vural, 2023), along with an academic accomplishment test as data collection instruments. The CTSAS scale is a 5-point Likert scale including 20 items and 5 sub-dimensions. In the reliability analysis of the scale, Cronbach's alpha value was .88, algorithmic thinking .81, decomposition .83, parallelization .88, abstraction .85, and automation .90. The Academic Achievement Test (AAT) comprises 25 multiple-choice questions, each with four answer alternatives. The questions of test were written in a clear expression and in a way that would not cause ambiguous understandings. The item difficulty of the test is 0.58, and the item discrimination rate is .50. The reliability test (Kuder Richardson KR-20) value was calculated as .89. According to these values, the test is of medium difficulty, medium selectivity and reliability.

Considering the sample size for the normality assumption, which is one of the general assumptions in statistical measurements, kurtosis and skewness values were considered in determining the normality distribution of AAT and CTSAS scores in repeated measurements. It was determined that AAT and CTSAS scores showed normal distribution. Repeated measurements from nonparametric tests were subjected

to the Friedman test due to the inability to guarantee the homogeneity of the dependent variable CTSAS scores and the equality of the variance and covariance matrices. To prevent Type 1 and Type 2 errors in repeated measurements, the MANCOVA test was used to analyze the significance level of the groups' AAT score differences. The multivariate statistical hypothetical criteria of MANCOVA were first checked. To check the linearity of the variables, it was tested whether the relationship between the dependent variables (post-test - retention scores) and the covariate variable (pre-test scores) in each group was linear or not and it was determined that there was a moderate relationship according to the regression lines. Also, Friedmann test, Wilcoxon test and Mann-Whitney U test were used in the study in cases where the assumptions of parametric tests were not met.

It was determined that the equality of covariance matrices ($\text{Box}_M = 4.96, F_{3-3012263} = 1.635, p = .179$) was ensured. Another assumption, the equality of error variances, was found to be met for the post-test scores ($F_{1-120} = .051, p = .821$) and retention-test ($F_{1-120} = 1.773, p = .186$) according to Levene's tests. Within the assumption of the equality of variances, the Mauchly sphericity test assumption ($W_0 = .985, \chi^2 = 1.83, p = .40$) was met with the data. Consequently, to fulfill these assumptions, the application of the MANCOVA method for repeated measures of the all tests scores of the academic achievement test was considered suitable.

Application

This study utilized the GBSL method, recognized as an innovative approach in coding education. This method aims to teach the intended objectives to learners by linking them to their real-life experiences and offering opportunities for engagement with professional groups in society. This study presents significant findings that contribute to the existing literature regarding the acquisition and assessment of CT skills, specifically through the use of the CT Skills Assessment Scale (CTSAS) in Scratch programming education.

The implementation was conducted over 13 weeks across three distinct secondary schools with varying socioeconomic statuses in the Milas area. Three weeks were allocated for the administration of pre-test, post-test, and retention tests, two weeks for the pilot study, and eight weeks for the intervention. The ICT course was two hours per week in all groups and four hours per week for each school type. The implementation was conducted by ICT educators in these institutions, with the researcher participating solely as an observer in the classroom setting. Before the application, the AAT and the CTSA Scale were administered to all groups within the same week across all three schools. Afterwards, a pilot study was conducted on the same groups for two weeks, and problems that were not taken into consideration or could not be predicted during the planning phase were observed and tried to be eliminated in this process. At the beginning of these problems, problems such as the way the implementing teachers presented the scenarios and the fact that the scenarios were not clear and understandable by the students were encountered. The

circumstances were evaluated, and the application was executed for 8 weeks. In the study, the exp. group of students was subjected to the scenario-based Scratch curriculum, whereas the control group utilised the pre-existing National Ministry of Education Program employed by the practitioner teachers in the prior procedure. A sample scenario applied to the experimental group is illustrated in Figure 1.

Role : Paediatrician

Scenario : You have encountered a case of poisoning while on duty in the emergency department. A 5-year-old child ate coloured painkiller pills at home, mistaking them for sugar. In the evening, he developed fever and was brought to the emergency department. With the treatment you provided, the child regained his former health. However, when you talked to his family about the negligence that caused this situation, you realised that his family was not very knowledgeable about the storage and control of medicines. For example; not putting the medicines in places that are not easily accessible, not paying attention to the expiry dates, in which conditions the medicines should be stored (in room conditions, out of the sun, refrigerator, etc.). When you encountered a few more similar cases, you realised that this was an important problem.

Mission : Utilizing the Scratch tool, you will create an animation illustrating the proper storage of medications. Medications are categorized into three types. Initially, the user's age will be obtained, and if the age exceeds 18, the application will indicate the appropriate storage conditions for the medicine based on its type. If the drug belongs to the Antibiotic category, the program will output 'refrigerator'; if it is a painkiller, 'in room conditions'; and if it is an antipyretic, 'out of the sun'. The application will be terminated if the user is under 18 years of age. Furthermore, the background (decor) or dummy will exhibit varying appearances based on the response.

Figure 1. Sample of experimental group's scenario

Within the scope of the MoNE program, teachers used unplugged coding activities and Scratch tools in the processing of the unit. The main difference in this program was that the teachers directly presented the Scratch programs they developed to the students and applied them through demonstration. Students were asked to develop these programs on their computers, sometimes individually and sometimes in groups. Q&A activities, demonstration, and explanation techniques were mainly used.

In addition to the researcher, two faculty members, three information technologies teachers, and a curriculum development specialist actively participated in the

development process of the scenario-based Scratch curriculum. A total of eight scenarios were developed within the scope of the curriculum. Lesson plans were designed to cover the learning outcomes targeted by these scenarios and to describe the teaching situations. At the end of the 8-week trial application, the post-test for AAT and CTSAS was administered. Retention test was applied to the groups six weeks following the post-test.

Findings

Findings Related to CTSAS Scores of Experimental and Control Group

Table 2 presents the descriptive statistical data for the total scores of the exp. and control groups from the pre-test, post-test, and retention-test collected from the CTSAS.

Table 2.

Total Scores of CTSAS of Groups

| Groups | Measure | N | Mean | sd | Min. | Max. |
|---------------------------|----------------|----|-------|-------|-------|--------|
| Experimental Group | Pre Test | 63 | 66.88 | 13.15 | 38.00 | 92.00 |
| | Post Test | 63 | 74.90 | 10.98 | 43.00 | 96.00 |
| | Retention-test | 63 | 74.19 | 11.28 | 52.00 | 100.00 |
| Control Group | Pre Test | 59 | 69.35 | 14.91 | 24.00 | 100.00 |
| | Post Test | 59 | 72.38 | 16.23 | 25.00 | 96.00 |
| | Retention-test | 59 | 70.55 | 14.40 | 23.00 | 96.00 |

According to Table 2, the mean pre-test total scores of the experimental group students ($\bar{X} = 66.88$, $sd = 13.15$) were lower than the mean scores of the control group ($\bar{X} = 69.35$, $sd = 14.91$), the mean post-test ($\bar{X} = 74.9$, $sd = 10.98$) and retention-test ($\bar{X} = 74.19$, $sd = 11.27$) total scores of the exp. group were higher than the mean of the post-test ($\bar{X} = 72.38$, $sd = 16.23$) and retention-test ($\bar{X} = 70.55$, $sd = 14.4$) total test scores of the control group. The findings related to the Friedman test applied to determine the sig. diff. in the test scores were shown in Table 3.

Table 3.

Friedmann Test and Wilcoxon Test Results of Experimental and Control Groups

| | N | Friedman χ^2 | sd | p | Wilcoxon Z | Cor. p (0.0167) | |
|--------------------------|----|----------------------|----|------|----------------------|--------------------|-------|
| Exp. Group | 63 | 19.56 | 2 | .000 | Pre test-Post test | 3.57 | .000* |
| | | | | | Post test- Ret. Test | .168 | .866 |
| | | | | | Pre test- Ret. Test | .001 | .001* |
| Control Group | 59 | 1.43 | 2 | .487 | Pre test-Post test | 1.034 | .301 |
| | | | | | Post test-Ret. Test | .415 | .678 |
| | | | | | Pre test- Ret. Test | -.464 | .642 |

While a sig. diff. was measured between the pre test - post test scores and pre test - retention test scores of the exp. group ($p < .005$), there was no sig. diff. between the

test scores of the control group. To ascertain the specific test scores that contributed to the difference in the exp. group's scores, the Wilcoxon Z comparison test was employed, utilizing the Bonferroni correction procedure to establish the adjusted significant value ($p=0.05/3=0.0167$). The analysis revealed a sig. diff. in the post-test total scores of the exp. group and the pre-test total scores, as well as between the pre-test total scores and the retention-test scores ($p<.0167$); however, no sig. diff. was found between the post-test and retention-test scores.

Is there a sig. diff. in favor of the Exp. Group when the total pre-test, post-test, and retention-test scores of the Exp. Group and the Control Group are compared in terms of the sub-dimensions (Algorithmic Thinking, Parallelisation, Decomposition, Abstraction, and Automation)?

Non-parametric, the Mann-Whitney U test was performed to ascertain if a significant difference existed among the pre-test, post-test, and retention-test scores derived from the sub-dimensions of the CTSAS between the exp. group students who experienced the scenario-based Scratch curriculum and the control group students. The analysis findings were showed in Table 4.

Table 4.
Results of the Mann-Whitney U Test for All Test Scores of CTSAS Factors

| Test | Groups | N | Mean of the Ranks. | Sum of the Ranks | U | W | Z | p | |
|----------------------|-----------------|---------------|--------------------|------------------|---------|--------|--------|--------|-------|
| Algorithmic Thinking | Pre test | Experimental | 63 | 56.56 | 3653 | 1547 | 3563 | -1.60 | .110 |
| | | Control | 59 | 66.78 | 3940 | | | | |
| | Post Test | Experimental. | 63 | 61.56 | 3878.50 | 1789 | 3559 | -.021 | .984 |
| | | Control | 59 | 61.43 | 3624.50 | | | | |
| | Retention Testi | Experimental. | 63 | 62.00 | 3906 | 1645 | 3415 | -.162 | .871 |
| | | Control | 59 | 60.97 | 3597 | | | | |
| Parallelisation | Pre test | Experimental | 63 | 60.02 | 3781 | 1765 | 3781 | -.48 | .63 |
| | | Control | 59 | 63.08 | 3772 | | | | |
| | Post Test | Experimental. | 63 | 62.53 | 3939 | 1793.5 | 3563 | -.33 | .74 |
| | | Control | 59 | 60.40 | 3563 | | | | |
| | Retention-testi | Experimental. | 63 | 63.98 | 4031 | 1702 | 3472 | -.80 | .42 |
| | | Control | 59 | 58.85 | 3472 | | | | |
| Decompositio n | Pre test | Experimental | 63 | 59.60 | 3754 | 1738.5 | 3754 | -.619 | .536 |
| | | Control | 59 | 63.53 | 3748 | | | | |
| | Post Test | Experimental. | 63 | 62.17 | 3916 | 1816.5 | 3586 | -.216 | .829 |
| | | Control | 59 | 60.79 | 3586 | | | | |
| | Retention-testi | Experimental. | 63 | 64.63 | 4072 | 1661 | 3431 | -1.018 | .309 |
| | | Control | 59 | 58.15 | 3431 | | | | |
| Abstraction | Pre test | Experimental | 63 | 60.07 | 3784.5 | 1768.5 | 3784.5 | -.466 | .641 |
| | | Control | 59 | 63.03 | 3718.5 | | | | |
| | Post Test | Experimental. | 63 | 68.80 | 4334.5 | 1398.5 | 3168.5 | -2.385 | .017* |
| | | Control | 59 | 53.70 | 3168.5 | | | | |
| | Retention Testi | Experimental. | 63 | 71.53 | 4506.5 | 1226.5 | 2996.5 | -3.268 | .001* |
| | | Control | 59 | 50.79 | 2996.5 | | | | |
| Auto mati on | Pre test | Experimental | 63 | 58.6 | 3692 | 1676 | 3692 | -.941 | .347 |
| | | Control | 59 | 64.59 | 3811 | | | | |
| Post Test | Experimental. | 63 | 60.73 | 3826 | 1810 | 3826 | -.25 | .802 | |

| | | | | | | | |
|------------------------|---------------|----|-------|---------|---------|---------|-------------|
| | Control | 59 | 62.32 | 3677 | | | |
| Retention Testi | Experimental. | 63 | 60.07 | 3784.5 | -1768.5 | 3784.5- | .46 .642 |
| | Control | 59 | 63.3 | 3718.5 | | | |
| Pre test | Experimental | 63 | 57.98 | 3653.00 | -1637 | 3653 | -1.135. 256 |
| | Control | 59 | 65.25 | 3850.00 | | | |
| Post Test | Experimental. | 63 | 62.60 | 3944.00 | -1789 | 3559 | -.356 .722 |
| | Control | 59 | 60.32 | 3559.00 | | | |
| Retention Testi | Experimental. | 63 | 64.89 | 4088.00 | -1645 | 3415 | -1.095.274 |
| | Control | 59 | 57.88 | 3415 | | | |

The analyses indicated no sig. diff. in the exp. group and the control group concerning tests' scores ($U_{\text{pre-test}} = 1637, p > .005$; $U_{\text{post-test}} = 1789, p > .005$; $U_{\text{Ret-test}} = 1645, p > .005$). Analysis of the rank averages and sums revealed that the exp. group students had a mean rank of 57.98 and a total rank sum of 3653 for their pre-test scores. In contrast, the control group students exhibited a mean rank of 65.25 and a total rank sum of 3850 for their pre-test scores. The rank averages indicated no sig. diff. between the groups despite the exp. group exhibiting lower values than the control group in pre-test rank scores and higher values in post-test and retention-test rank scores.

The tests' scores of the algorithmic thinking sub-dimension of the CTSAS yielded no sig. diff. in the exp. group and the control group ($U_{\text{pre-test}} = 1547, p > .005$; $U_{\text{post-test}} = 1789, p > .005$; $U_{\text{ret.-test}} = 1645, p > .005$). The exp. group's pre-test scores were 56.56 and 3653, respectively, when the rank averages and sums were analyzed. The pre-test scores of the control group were 66.78 and 3940, respectively. Although the exp. group had a lower value than the control group in the pre-test rank scores, the exp. group had a higher value in the post-test and retention-test rank scores. The rank averages revealed that there was no sig. diff. in the groups. However, the exp. group had a higher value in the retention-test rank scores.

No sig. diff. were observed in the all tests' scores for the parallelization sub-dimension ($U_{\text{pre-test}} = 1765, p > .05$; $U_{\text{post-test}} = 1793.5, p > .05$; $U_{\text{Ret-test}} = 1702, p > .05$). Analysis of the rank averages and sums revealed that the exp. group students had a mean rank of 60.02 and a total rank sum of 3781 for their pre-test scores. In contrast, the control group students exhibited a mean rank of 63.08 and a total rank sum of 3772 for their pre-test scores. The rank averages indicated no sig. diff. in the groups despite the exp. group exhibiting lower pre-test rank scores compared to the control group and higher post-test and retention-test rank scores.

All tests' scores of the decomposition sub-dimension were not sig. diff. ($U_{\text{pre test}} = 1738.5, p > .005$; $U_{\text{post test}} = 1816.5, p > .05$; $U_{\text{Ret.-test}} = 1661, p > .05$). When the rank averages and sums were analyzed, the mean and sum of the ranks of the exp. group students' pre-test scores were 59.60 and 3754, respectively, while the mean and sum of the ranks of the control group's pre-test scores were 66.53 and 3748, respectively. According to the rank averages, although the exp. group had a lower value than the control group in the pre-test rank scores and a higher value in the post-test and retention-test rank scores, difference was not significant ($p > .05$).

Upon analyzing the rank averages and sums of the abstraction sub-dimension, the exp. group showed a mean rank of 60.07 and a rank sum of 3784.5 in the pre-test scores, while the control group revealed a mean rank of 63.03 and a rank sum of 3718.5 in the pre-test scores. The post-test scores indicated that the exp. group had a mean rank of 68.80 and a total rank score of 4334.5, whereas the control group had a mean rank of 53.70 and a total rank score of 3168.5. The exp. group had a mean rank of 71.53 and a total rank score of 4506.5 in the retention exam, whereas the control group had a mean rank of 50.79 and a total rank score of 2996.5. The rank averages indicated that the exp. group exhibited lower pre-test rank scores compared to the control group, but higher post-test and retention-test rank scores. The analysis revealed no sig. diff. in the groups at the abstraction sub-dimension of the CTSA scale pre-test scores; however, a sig. diff. favoring the exp. group was observed in the post-test and retention-test scores. ($U_{Pre-test} = 1768.5, p > .05$; $U_{Post-test} = 1398.5, p < .05$; $U_{Ret-test} = 1226.5, p < .05$).

Upon analyzing the rank averages and total rank scores within the automation sub-dimension, the exp. group students exhibited mean and total rank scores of 58.6 and 3692 in the pre-test, whereas the control group demonstrated mean and total rank scores of 64.59 and 3811 in the pre-test. The analysis revealed no significant change in the all tests' scores of the automation sub-dimension of the CTSAS between the exp. and control groups. ($U_{Pre-test} = 1676, p > .05$; $U_{Post-test} = 1810, p > .05$; $U_{Retention} = 1768.5, p > .05$).

Is there a sig. diff. in the Exp. Group and the Control Group in the CTSAS Sub-dimensions: Tests's Total Scores?

Table 5 presented the results of the Friedman test applied to evaluate the differences among the test scores.

Table 5.

Friedman Test Results Utilized to Compare Scores

| Sub-Dimensions | Groups | N | Friedman (χ^2) | sd | p | Wilcoxon Z | Corr. p (0.0167) | |
|----------------------|--------------------|----|-----------------------|----|-------|---------------------|------------------|--------|
| Algorithmic Thinking | Experimental Group | 63 | 18.42 | 2 | .000* | Pre test-Post test | 3.42 | .001** |
| | | | | | | Post test-Retention | .771 | .441 |
| | | | | | | Pre test- Retention | -2.87 | .004** |
| | Control Group | 59 | .078 | 2 | .962 | | | |
| Paralelisation | Experimental Group | 63 | 5.51 | 2 | .063 | | | |
| | | | | | | Control Group | 59 | .22 |
| Decomposition | Experimental Group | 63 | 13.209 | 2 | .001* | | | |
| | | | | | | Control Group | 59 | .670 |
| Abstraction | Experimental Group | 63 | 15.078 | 2 | .001* | Pre test-Post test | 2.66 | .008** |
| | | | | | | Post test-Ret. test | .637 | .524 |
| | | | | | | Pre test- Ret. test | -1.72 | .084 |
| | Control Group | 59 | .694 | 2 | .707 | | | |

| | | | | | | | |
|------------|---------------|-------|---|-------|---------------------|-------|--------|
| Automation | Experimental | 6.214 | 2 | .045* | Pre test-Post test | 2.617 | .009** |
| | Group | | | | Post test-Ret. test | -.216 | .829 |
| | | | | | Pre test- Ret. test | 2.59 | .009** |
| | Control Group | 1.929 | 2 | .381 | | | |

The examination of the algorithmic thinking sub-dimension indicated a sig. diff. in the total scores of the exp. group in the all tests' scores. The control group demonstrated no sig. diff. in total scores across these measurements. The Bonferroni correction method was employed to identify the differences in test scores within the exp. group. The adjusted significance value was calculated ($p=0.05/3=0.0167$), and the Wilcoxon Z comparison test was subsequently applied. The analysis revealed a sig. diff. in the total pre-test scores of the exp. group and the total post-test and retention-test scores ($p<.0167$). However, no sig. diff. was found between the post-test and retention-test scores. ($p>.0167$).

The analysis of the parallelization sub-dimension revealed no sig. diff. in the pre-test, post-test, and retention-test scores between groups. ($p>.05$).

The exp. group shown a sig. diff. in total scores of all tests in the decomposition sub-dimension. In contrast, the control group showed no sig. diff. in total scores between the pre-test, post-test, and retention-test. The Wilcoxon comparison test, supplemented by the Bonferroni correction method, was employed to identify the differences in test scores among the exp. group. The analysis results indicate a significant difference in the post-test total scores of the experimental group and the pre-test total scores ($p < .0167$). However, no sig. diff. was found between the pre-test total scores and the retention-test scores, nor between the post-test and retention-test scores. ($p>.0167$).

The Friedman and Wilcoxon Z Test results for the Abstraction sub-dimension indicated a sig. diff. in tests' total scores of the exp. group. In contrast, no sig. diff. was observed in the total scores of the control group. The Wilcoxon comparison test, supplemented by the Bonferroni correction method, was utilized to identify the differences in test scores among the exp. group. The analysis revealed a sig. diff. in the post-test total scores of the exp. group and the pre-test total scores ($Z=2.97$; $p<.0167$), as well as between the post-test and retention-test scores ($Z=-3.76$; $p<.0167$). However, no sig. diff. was found between the pre-test and the retention-test scores ($p>.0167$). No sig. diff. was observed in the all tests' scores of the exp. group ($p > .05$) and those of the control group.

Findings of the Scenario-Based Curriculum on Students' AAT scores

Descriptive statistics and repeated measures analysis of covariance method (MANCOVA) were used to examine the research question: was there a sig. diff. in the mean scores of the post-test and retention-test in favor of the exp. group when the total mean scores of the pre-test of the AAT of the exp. group students to whom the scenario-based Scratch curriculum was applied and the control group students to

whom the MoNE curriculum was applied. The descriptive statistical results of the mean scores of the all tests obtained in the AAT of the groups were shown in Table 6.

Table 6.
Descriptive Statistical Results of Mean Scores of AAT

| Tests | Groups | \bar{X} | sd | Y_i | Std. Err. | 95% Confidence Interval | | Y_j | Std. Err. | 95% Confidence Interval | |
|----------------|--------------|-----------|-------|-------|-----------|-------------------------|-------------|-------|-----------|-------------------------|-------------|
| | | | | | | Lower Bound | Upper Bound | | | Lower Bound | Upper Bound |
| Pre Test | Experimental | 43.52 | 18.98 | | | | | | | | |
| | Control | 49.24 | 16.89 | | | | | | | | |
| | Total | 46.29 | 18.16 | | | | | | | | |
| Post Test | Experimental | 64.92 | 19.79 | 67.1 | 1.69 | 63.71 | 70.43 | | | | |
| | Control | 62.79 | 18.86 | 60.49 | 1.75 | 57.02 | 63.96 | | | | |
| | Total | 63.89 | 19.30 | 63.78 | 1.22 | 61.38 | 66.18 | | | | |
| Retention-test | Experimental | 69.63 | 18.47 | 71.63 | 1.86 | 67.94 | 75.31 | 69.35 | 1.50 | 66.37 | 72.33 |
| | Control | 65.00 | 20.68 | 62.87 | 1.92 | 59.05 | 66.68 | 61.68 | 1.55 | 58.60 | 64.76 |
| | Total | 67.39 | 19.62 | 67.25 | 1.33 | 64.61 | 69.88 | | | | |

To examine the AAT scores of the groups, the post-test, retention-test scores, and mean scores of the groups were checked. Accordingly, it was seen that the post-test and retention-test adjusted mean scores of the exp. group ($Y_j=69.35$, $sd=1.5$) were higher than the post-test and retention-test-adjusted mean scores of the students in the control group ($Y_j=61.68$, $sd=1.55$). Within the groups, when the pre-test scores of the exp. group were controlled, the post-test scores of exp. group were $Y_i=67.1$, $sd=1.69$, and the retention-test scores were $Y_i=71.6$, $sd=1.86$, while the average post-test scores of the control group were $Y_i=60.49$, $sd=1.75$ and the retention-test scores were $Y_i=62.87$, $sd=1.92$.

The exp. group exhibited lower pre-test scores; however, a greater increase in scores was noted in the post-test results for this group. Additionally, while both groups showed increases in retention-test scores, the exp. group outperformed the control group in both post-test and retention-test scores.

The results of MANCOVA for repeated measures of the mean scores of the academic achievement post-test and retention-test for the groups, with academic achievement pre-test scores controlled, were showed in Table 7.

Table 7.
MANCOVA Results for Repeated Measures of AAT Scores

| Source of Variance | Sum of Squares | Sd | Mean Square | F | p | η^2 |
|--------------------|----------------|----|-------------|---|---|----------|
|--------------------|----------------|----|-------------|---|---|----------|

Within Groups

| | | | | | | |
|---|-----------|-----|----------|--------|-------|------|
| Model | 29788.25 | 1 | 29788.25 | 105.61 | .000* | .47 |
| Pre Test | 43867.90 | 1 | 43867.90 | 155.53 | .000* | .57 |
| Groups (Experimental-Control) | 3495.46 | 1 | 3495.46 | 12.93 | .001* | .09 |
| Error | 33563.65 | 119 | 282,048 | | | |
| Between Groups | | | | | | |
| Measure (Pre Test-Post Test- Retention-test) | 296.557 | 1 | 296.557 | 2.638 | .107 | .022 |
| Measure * Pre test | 64.876 | 1 | 64.876 | .577 | .449 | .005 |
| Measure * Group | 70.648 | 1 | 70.648 | .629 | .429 | .005 |
| Error | 13376.332 | 119 | 112.406 | | | |

The analysis of repeated measures MANCOVA indicated that pre-test scores significantly predicted the post-test and retention-test scores of all groups. ($F_{1-119}=155.53, p=.00$). A statistically sig. diff. in post-test and retention-test scores was found in the exp. and control groups according to the results of the repeated measures test. ($F_{1-119}=12.93, p=.001, \eta^2=.09$). When the results within groups were analysed, it was determined that the diff. in all tests' scores in the measurement was not significant ($F_{1-119}=2.64, p=.107$), and the interaction of the measurements with the groups was not significant ($F_{1-119}=.629, p=.429$). While it was seen that the results of the group averages were within the 95% confidence interval, it was seen that the difference in the groups was significant in favor of the experimental group ($F_{1-119}=12.93, p=.001, \eta^2=.09$) and that this experimental application was effective in increasing academic achievement. The experimental application was found to have a 9% effect on the difference in the groups. The partial Eta-Square value ($\eta_{kismi}^2=.09$) was taken as the basis for the calculation of this effect power.

Analysis of Post-Test and Retention-test Scores of Students by Gender, Controlling for AAT Pre-Test Total Scores

Analysis of Table 8 revealed that, when controlling for pre-test scores, AAT pre-test scores significantly predicted academic achievement in both post-test and retention-test scores ($F_{1-117}= 108.001, p=0.00$). The repeated measures test conducted between the exp. and control groups regarding the gender variable revealed a significant difference in post-test and retention-test mean scores in the groups ($F_{1-117}=4.99, p=.003, \eta^2=.09$).

Table 8.

Repeated Measures MANCOVA Results of AAT Scores in Terms of Gender

| Source of Variance | Sum of Squares | Sd | Mean Square F | p | η^2 |
|---------------------------|-----------------------|-----------|----------------------|----------|----------------------------|
| Between Groups | | | | | |
| Model | 15162.63 | 1 | 15162.639 | 108.001 | .000* |
| Pre Test | 20334.46 | 1 | 20334.467 | 144.839 | .000* |
| Group * Gender | 2103.55 | 3 | 701.186 | 4.994 | .003* |
| Error | 16426.001 | 117 | 140.393 | | |
| Within Groups | | | | | |

| | | | | | | |
|------------------------------------|-----------|-----|---------|-------|------|------|
| Measure (Pre-Post-Retention-tests) | 467.056 | 1 | 467.056 | 2.092 | .151 | .018 |
| Measure * Pre test | 57.299 | 1 | 57.299 | .257 | .613 | .002 |
| Measure*Group*Gender | 778.769 | 3 | 259.590 | 1.163 | .327 | .029 |
| Error | 26115.191 | 117 | 223.207 | | | |

The contrast analysis technique was utilized to identify the groups exhibiting differences, which are detailed in Table 9.

Table 9.
Between Groups Contrast Analysis Estimation (CAE)

| | | 95% Confidence Interval for Difference | | | | |
|----------------------------|-------------------|--|----------|-------|-------------|-------------|
| (I) Gender Groups | (J) Gender Groups | Mean Differences (I-J) | Std Err. | p | Lower Bound | Upper Bound |
| Experimental-Female | Exp.-Male | 3.838 | 2.996 | 1.000 | -4.203 | 11.880 |
| | Control-Female | 11.164* | 3.142 | .003 | 2.731 | 19.598 |
| | Control-Male | 8.121* | 2.935 | .039 | .244 | 15.998 |
| Experimental-Male | Exp.-Female | -3.838 | 2.996 | 1.000 | -11.880 | 4.203 |
| | Control-Female | 7.326 | 3.212 | .146 | -1.295 | 15.946 |
| | Control-Male | 4.283 | 3.047 | .975 | -3.895 | 12.460 |
| Control -Female | Exp.-Female | -11.164* | 3.142 | .003 | -19.598 | -2.731 |
| | Exp.-Male | -7.326 | 3.212 | .146 | -15.946 | 1.295 |
| | Control -Male | -3.043 | 3.169 | 1.000 | -11.547 | 5.461 |
| Control-Male | Exp.-Female | -8.121* | 2.935 | .039 | -15.998 | -.244 |
| | Exp.-Male | -4.283 | 3.047 | .975 | -12.460 | 3.895 |
| | Control-Female | 3.043 | 3.169 | 1.000 | -5.461 | 11.547 |

The contrast analysis results indicated no sig. diff. in the post-test and retention-test mean scores of female students and male students in the exp. group (CAE = 3.83, sd = 2.99, p = 1.0). Contrast analysis indicated a sig. diff. in post and retention test mean scores between female students in the exp. group and both male and female students in the control group, favoring the female students in the exp. group (control-female group: CAE=11.16, sd=3.14, p<.05; control-male group: CAE=8.12, sd=2.93, p<.05). No sig. diff. was observed between the male students in the exp. group and the male and female students in the control group regarding the mean scores of the post-test and retention-test (control-female group: CAE = 7.32, sd = 3.21, p = .146; control-male group: CAE = 4.28, sd = 3.04, p = .975). The intervention was found to have an 11% effect on the difference in the groups. The effect size calculation was based on the partial Eta-Squared value. ($\eta_{kismi}^2 = .11$).

Post-Test and Retention-test Scores of Groups by School Type when AAT Pre-Test Total Scores were Controlled

The analysis of repeated measures MANCOVA, conducted under controlled pre-test scores, revealed that the pre-test scores significantly predicted the post-test and retention-test scores for both the exp. and control groups, categorized by school type. Furthermore, the pre-test scores, regarded as covariate variables, exhibited a 57% interaction effect on the post-test and retention-test scores. ($F_{1-115}=152.005$, $p=0.00$, $\eta^2=.57$). Table 10 shows the findings of the repeated measures MANCOVA results for the examination of the mean scores of the AAT in terms of school type.

Table 10.

Repeated Measures MANCOVA Results of AAT Scores in Terms of School Type

| Source of Variance | Sum of Squares | Sd | Mean Square | F | p | η^2 |
|--------------------------------|----------------|-----|-------------|---------|-------|----------|
| Between Groups | | | | | | |
| Measure (Post- Pre Test) | 27269.093 | 1 | 27269.093 | 100.129 | .000* | .465 |
| Pre Test | 41397.016 | 1 | 41397.016 | 152.005 | .000* | .569 |
| School type | 144.045 | 2 | 72.022 | .264 | .768 | .005 |
| Group | 4224.023 | 1 | 4224.03 | 15.510 | .000* | .119 |
| School type* Group | 2096.547 | 2 | 1048.273 | 3.849 | .024* | .063 |
| Hata | 31319.115 | 115 | 272.340 | | | |
| Within Group | | | | | | |
| Measure (Post- Retention-test) | 250.694 | 1 | 250.694 | 2.390 | .125 | .020 |
| Measuser * Pre test | 80.189 | 1 | 80.189 | .764 | .384 | .007 |
| Measure * School type | 582.932 | 2 | 291.466 | 2.778 | .066 | .046 |
| Measure * Group | 290.898 | 1 | 290.898 | 2.773 | .099 | .024 |
| Measure * School type * Group | 737.468 | 2 | 368.734 | 3.515 | .033* | .058 |
| Error | 12065.153 | 115 | 104.914 | | | |

In the results regarding the effects of the scenario-based Scratch curriculum on the AAT scores within the groups; the difference in the post-test and retention-test scores was not significant ($F_{1-115}=2.39$, $p=.125$), the interaction of the measurements with the pre-test was not significant ($F_{1-115}=.764$, $p=.384$), the interaction between measurements and groups was not significant ($F_{1-115}=2.77$, $p=.09$), the interaction of measurements with school type was not significant ($F_{1-115}=2.78$, $p=.066$), and the interaction of measurements with school type and groups was significant ($F_{2-115}=3.51$, $p=.033$).

Measurements conducted across various school types revealed a significant difference in the experimental group, which utilized the Scenario-based Scratch curriculum, and the control group, which implemented the MoNE curriculum, regarding school type in post-test and retention-test scores following repeated assessments to ascertain the interaction between the experimental and control groups. ($F_{2-115}=3.85$, $p=.024$, $\eta^2=.06$). To examine the difference in the school types, contrast analysis was checked by using the post-test and retention-test mean scores adjusted according to the pre-test scores.

Table 11.
Between Groups Contrast Analysis Estimation (CAE)

| (I) School Type | (J) School Type | Mean Differences (I-J) | Sd. | p | 95% Confidence Interval Mean Differences | |
|----------------------|----------------------|------------------------|-------|-------|--|------------|
| | | | | | Min. Bound | Max. Bound |
| High Socioeconomic | Middle Socioeconomic | 4.878 | 3.230 | .134 | -1.520 | 11.275 |
| | Lower Socioeconomic | 6.523 | 3.429 | .030* | -.270 | 13.315 |
| Middle Socioeconomic | Higher Socioeconomic | 4.878 | 3.230 | .134 | -11.275 | 1.520 |
| | Lower Socioeconomic | 1.645 | 2.825 | .562 | -3.951 | 7.241 |
| Low Socioeconomic | Higher Socioeconomic | -6.523 | 3.429 | .030* | -13.315 | .270 |
| | Middle Socioeconomic | -1.645 | 2.825 | .562 | -7.241 | 3.951 |

According to the contrast analysis between the socioeconomic upper-level school type and the socioeconomic lower-level school type in the experimental group, it was determined that there was a significant difference in favor of the socioeconomic upper-level school type (CAE = 6.523 sd=3.429 p=.03).

While it was seen that the results of the group averages were within the 95% confidence interval, it was seen that there was a significant difference between the groups in favor of the socioeconomic upper-level school-type experimental group in the post-test scores ($F_{1-115}=15.51, p=.000, \eta^2=.12$) and that this experimental application was effective in increasing academic achievement. Accordingly, it was determined that the experimental application had an effect of 12% on the difference between the groups. The calculation of this effect power was based on the partial Eta-Square value ($\eta^2_{\text{partial}}=.12$).

Findings in Terms of Screen Use Between Post-Test and Retention-Test Scores When the Total Scores of the Groups AAT Pre-Test were Controlled

Table 12 showed the repeated measures MANCOVA outcomes for the post-test and retention-test scores of the groups, accounting for screen usage frequency while controlling for pre-test scores. The pre-test scores of AAT significantly predicted the post-test and retention-test scores, with the pre-test scores acting as a covariate variable, demonstrating a 47% interaction in the post-test and retention-test scores ($F_{1-113}= 98.772, p=0.00, \eta^2= .47$). The results concerning the impact of AAT scores within the exp. groups indicated that the difference in post-test and retention-test scores was not significant ($F_{1-113}=3.25, p=.074$). Additionally, the interaction of the measurements with the pre-test was not significant ($F_{1-113}=1.093, p=.298$), the interaction between the measurements and the groups was not significant ($F_{1-113}=.076, p=.784$), the interaction of the measurements with the frequency of screen use was not significant ($F_{1-113}=.478, p=.698$), and the interaction of the measurements with both the frequency of screen use and the groups was not significant ($F_{3-113}=.327, p= .806$).

Table 12.

Repeated Measures MANCOVA Results of AAT Scores in Terms of Frequency of Screen Use

| Source of Variance | Sum of Squares | Sd | Mean Square | F | p | η^2 |
|-------------------------------------|----------------|-----|-------------|---------|------|----------|
| Between Groups | | | | | | |
| Measure (Post - Retention-test) | 34009.317 | 1 | 34009.317 | 151.354 | .000 | .573 |
| Pre test | 22193.968 | 1 | 22193.968 | 98.772 | .000 | .466 |
| Group | 1213.367 | 1 | 1213.367 | 5.400 | .022 | .046 |
| Screen usage freq. | 7321.290 | 3 | 2440.430 | 10.861 | .000 | .224 |
| Screen usage freq. * Group | 760.146 | 3 | 253.382 | 1.128 | .341 | .029 |
| Error | 25391.094 | 113 | 224.700 | | | |
| Within Groups | | | | | | |
| Measure (Post - Retention-test) | 378.079 | 1 | 378.079 | 3.25 | .074 | .028 |
| Measure * Pre test | 126.78 | 1 | 126.78 | 1.093 | .298 | .010 |
| Measure * Screen usage freq. | 166.493 | 3 | 55.498 | .478 | .698 | .013 |
| Measure * Group | 8.787 | 1 | 8.787 | .076 | .784 | .001 |
| Measure * Screen usage freq.* Group | 113.670 | 3 | 37.89 | .327 | .806 | .009 |
| Error | 13107.45 | 113 | 115.995 | | | |

The measurements of screen use frequency revealed no significant difference between the exp. group, which utilized the GBSL Scratch curriculum, and the control group, which followed the MoNE curriculum, regarding post-test and retention-test scores. ($F_{3-113}=1.128, p=.341$). However, it was observed that there was a sig. diff. in the post-test and retention scores according to the frequency of screen use within the groups ($F_{3-113}=10.861, p=.000$). To examine the difference according to the frequency of screen use, contrast analysis was checked by using post and retention test mean scores corrected according to pre-test scores.

Table 13.

Contrast Analysis Scores of the Difference in Frequency of Screen Use

| (I) Screen usage freq. | (J) Screen usage freq. | Mean Differences (I-J) | Sd. | p | 95% Conf. Int. Mean Differences | |
|-------------------------|------------------------|------------------------|-------|-------|---------------------------------|------------|
| | | | | | Min Bound. | Max. Bound |
| 0-1 hour | 1-2 hours | .072 | 3.547 | 1.000 | -9.454 | 9.598 |
| | 2-3 hours | 9.685* | 3.606 | .050 | .000 | 19.370 |
| | 3 hours and more | 14.810* | 3.711 | .001* | 4.843 | 24.776 |
| 1-2 hours | 1-2 hours | -.072 | 3.547 | 1.000 | -9.598 | 9.454 |
| | 2-3 hours | 9.613* | 2.652 | .003* | 2.492 | 16.734 |
| | 3 hours and more | 14.737* | 2.782 | .000* | 7.265 | 22.209 |
| 2-3 hours | 1-2 hours | -9.685* | 3.606 | .050 | -19.370 | .000 |
| | 2-3 hours | -9.613* | 2.652 | .003* | -16.734 | -2.492 |
| | 3 hours and more | 5.124 | 2.445 | .230 | -1.442 | 11.690 |
| 3 hours and more | 1-2 hours | -14.810* | 3.711 | .001* | -24.776 | -4.843 |
| | 2-3 hours | -14.737* | 2.782 | .000* | -22.209 | -7.265 |

| | | | | | |
|------------------|--------|-------|------|---------|-------|
| 3 hours and more | -5.124 | 2.445 | .230 | -11.690 | 1.442 |
|------------------|--------|-------|------|---------|-------|

When Table 13. was examined; there was a sig. diff. between the groups with screen usage frequency of 0-1 hour and 1-2 hours and the groups with screen usage frequency of 2-3 hours and 3 hours or more in favor of 0-1 hour and 1-2 hour screen users (CAE= 14.810, sd= 3.71). When the difference between the mean scores of the post-test and retention-test for the duration of screen use was examined, there was a difference of (CAE=9.68, sd=3.606) between those who used between 0-1 hour and those who used between 2-3 hours, a difference of (CAE=14.81, sd=3.71) between those who used 0-1 hour and those who used 3 or more hours, while there was no significant difference between those who used between 0-1 hour and 1-2 hours (CAE=.072, sh=3.54). The measures revealed no sig. diff. in the Group * Screen Time interaction; however, it indicated that the mean scores of the post-test and retention-test varied based on the frequency of screen use, with the results favoring individuals with lower screen usage frequency. An effect of 22% was identified in the difference between the groups. The calculation of this effect size utilized the partial Eta-Square value ($\eta^2_{\text{partial}} = .224$) as a reference.

Conclusion and Suggestions

In the study, an AAT was applied before and after the implementation of the GBSL Scratch curriculum and 6 weeks after the end of the implementation. As a result of this application, important findings were obtained regarding the effectiveness of the program in the groups, differences in terms of gender, socioeconomic level, and frequency of screen use. In the study, it was observed that the GBSL created a significant difference in both the post-test and retention-test scores of the students, this difference was also examined in terms of gender, socioeconomic level, frequency of screen use, and differences were also observed according to these variables. While there are studies in previous studies (Karcı, 2018; Kemiksiz, 2016; Yaman, 2005; Schoenfeld-Tacher et al. 2001) that have a positive effect on academic achievement, fewer studies have found that there was no sig. diff. (Top&Arabacıoğlu, 2024).

The test to assess the sig. diff. in post-test and retention-test scores between the exp. group, which utilized the GBSL Scratch curriculum, and the control group, which employed the MoNE curriculum, revealed a sig. diff. favoring the exp. group. The intervention significantly enhanced the academic achievement scores of the students. The average scores demonstrated a sig. diff. favoring the exp. group in both the post-test and the retention-test. Furthermore, the mean scores of the exp. group in the retention-test indicated that this effect persisted in the retention assessment. No sig. diff. was observed in the post-test and follow-up test scores of the groups when the pre-test scores were controlled in the in-group measurements. Upon controlling for pre-test scores, it was observed that there was no sig. diff. in the post-test and retention-test scores among the groups, resulting in a more reliable measurement by mitigating the effects and error rates associated with the pre-test through MANCOVA. Although the retention-test scores were anticipated to be inferior to the post-test scores, the average retention-test score in the study was higher. This is due to the

likelihood that students may have formed a habit and immunity to the test questions because the same test was utilized as a pre-test, post-test, and retention-test.

Research (Top & Arabacıoğlu, 2024; Hursen & Fıslı, 2017; Green et al., 2017; Kemiksiz, 2016; Samsa et al., 2009; Bell, 1994) indicates a sig. diff. in post-test and retention-test scores between groups (Yeniceli, 2016; Ciraj et al., 2010) following the implementation of the GBSL. Kemiksiz (2016)'s investigation showed the impact of GBSL on academic performance in science courses and asserted that the program developed with GBSL was more effective than the instruction provided under the MoNE curriculum, highlighting a sig. diff. in both post-test and retention-test scores between the groups. Hursen and Fıslı (2017) discovered that the academic performance of pre-service teachers exceeded that of the control group, which engaged in reflective thinking activities, following a 12-week GBSL training program for pre-service teachers. Gossman et al. (2007) asserted that real-life scenarios are an effective pedagogical approach for students, enhancing retention in learning. Özsevgeç and Kocadağ (2013) arrived at analogous conclusions in their research on prospective educators, indicating that the information acquired through this method was more enduring. Due to the constructivist approach at the basis of GBSL; individuals are more active in the learning process with GBSL and can use the knowledge they create in their minds more permanently. In GBSL environments, the teacher is in the position of constructing and constructing problems and putting students into roles from life, rather than transferring information and presenting information ready-made. In this way, instead of directly obtaining information, learners create it by themselves at the end of different thinking processes. In this process, it is known that learning is fully internalized and learning products are acquired more permanently. For this reason, active participation of learners in the learning process can be considered as an important parameter in explaining the increase in academic achievement, which has a conceptually broad definition and depends on many different variables.

The secondary sub-problem of the research investigates the sig. diff. between the post-test and retention-test scores of the groups, considering gender, while controlling for the pre-test total scores of the AAT. The research findings indicate that the post-test and retention-test scores of female students in the exp. group surpassed those of other groups, whereas the post-test scores of female students in the control group were the lowest among all groups. The exp. group exhibited the greatest diff. in the retention-test and post-test mean scores of female students, while the retention-test scores of the general groups improved relative to the post-test scores. In contrast, the retention-test and post-test scores of the control-male group were nearly the same. Consequently, it was determined that the disparity in measurements between the groups of the intervention program favoured the female students of the exp. group. Schoenfeld et al. (2001) asserted in their study on the influence of various factors in GBSL that GBSL provides equitable opportunities for learners across gender and ethnic groups. The Problem Solving and Programming unit includes coding and

programming topics, based on the premise that male students demonstrate a higher propensity for engagement and utilization of computers and programming. (Du et al., 2016; Lockheed, 1985). In recent years, it has been noted that female students have become increasingly active in programming and coding, a finding verified by the study (IEA., 2014; Colley et al., 2003). Students' comprehension of programming structures did not differ by gender in elementary programming instruction using the Scratch tool; however, female students demonstrated a superior understanding of algorithmic concepts. (Hsu, 2014).

A difference was identified between the mean scores of the retention test and post-test among groups in schools with varying socioeconomic statuses. The differences were identified between the upper-level socioeconomic school and the lower socioeconomic school, whereas the middle-level socioeconomic school exhibited no significant difference between the upper and lower socioeconomic schools. Due to the high-class size of the students in both the control and experimental groups and the hardware deficiencies, it was determined that not every student in the group could contribute to both the lesson and the in-class applications at the same level when using computers with the group. It was seen that the inadequate material and hardware of the classroom environment caused limitations in the application of the GBSL. Problem solving and programming unit of 6th grade Computer Technologies course, which is the unit covered in the application, has been frequently mentioned under the title of coding in recent years and many studies have been conducted on it (Kalelioğlu & Gülbahar, 2015; Karabak & Güneş, 2013; Brennan & Resnick, 2012; Stolee & Fristoe, 2011; Kelleher et al., 2007). For this reason, in the findings of the studies conducted in the field, there are many suggestions about the appropriate number of available equipment in the classroom (Yıldız Durak, 2018; Kelleher et al., 2007). The fact that high-level socioeconomic school students make continuous repetitions at home shows that programming subjects have an effective result in learning (Cevahir et al., 2017; Özmen et al., 2014; Lahtinen et al., 2005).

The measurements conducted regarding screen usage time revealed that there was no sig. diff. in the exp. group, which engaged with the GBSL Scratch curriculum, and the control group, which followed the MoNE curriculum, in terms of screen usage time when comparing post-test and retention-test scores. The analysis revealed that, although there was no sig. diff. concerning the interaction of group screen time, the mean scores of the post-test and retention-test exhibited a difference favoring individuals who engaged in less screen time. The data indicated a decline in academic achievement scores among the groups as the duration of screen usage among students escalated. Arabacıoğlu (2012) investigated the prevalence of internet usage and group interaction among students in two distinct communication settings structured with GBSL activities, ultimately concluding that the impact was not significant across the scale scores.

A sig. diff. was observed in the post-test and retention-test scores of the exp. group when evaluating the effect of the GBSL on the CT skills of students, while no sig. diff. was found in the control group. Analysis of the notable disparities between the groups revealed no sig. diff. between the exp. and control groups, even though the exp. group demonstrated higher mean scores in the pre-test, post-test, and retention assessments. A notable difference was observed between the exp. and control groups solely in the abstracting dimension of the scale's sub-dimensions.

Thus, although a sig. diff. between the groups was observed only in the abstraction dimension of the CTSAS, sign. diff. were evident in the pre-test, post-test, and retention-test scores of the exp. group across the other dimensions. In this context, while the conceptual clarity of the CT skill associated with this method remains incomplete, it is predominantly regarded as a problem-solving and cognitive process in the literature (Kop&Arabacıoğlu, 2024; Voskoglou et al., 2012; Csizmadia et al., 2015; Aho, 2012; Wing, 2011). GBSL substantially impacts the cultivation of CT skills. The lack of a comparable difference in the control group suggests that the results in the problem-solving and programming unit are due to the impact of GBSL, notwithstanding its importance in developing CT skills. Furthermore, specific studies have revealed a substantial correlation between programming proficiency and skills of CT. (Atiker, 2019). The GBSL approach encourages individuals to engage in higher-level thinking to accomplish classroom learning tasks, potentially leading to a shared area of growth in the development of critical thinking skills. No studies were identified in the literature review regarding CT with GBSL.

Abstraction is the analysis of problem situations by separating a problem into its components, without seeing unimportant details (Csizmadia et al. 2015; Saeli et al., 2005). Çetin (2017) stated that abstraction overlaps with Piaget's empirical abstraction, but it does not fully meet the meaning of learning or creating computer science concepts. Aharoni (2000) stated that there are three levels of abstraction in programming, the first level is programming language-based, the second level is programming-based and the third level is programming-independent. According to Aharoni (2000), at the first level, students who start programming solve problems by learning a specific programming language. At the second level, students do not need any programming language, they can perform platform-independent analysis, while at the last level, they can perform problem-solving without the need for any programming concepts (Çetin & Toluk-Uçar, 2019). Kukul (2018) concluded that there was a sig. diff. in the abstraction and decomposition dimensions of the students in the exp. group in the self-efficacy sub-scores when evaluating different programming training. It is understood from the definition that there is a close relationship between the use of the Scratch visual tool used in this study and students' abstraction skills. In recent years, many studies have been encountered in the literature on the effect of the Scratch program on CT skills (Yünkül et al., 2017; Calao et al., 2015). There were studies on the use of the Scratch tool and its relationship with CT

and its sub-skill areas of algorithmic thinking, problem-solving, and creativity (Çatlak et al., 2015; Olabe et al., 2011).

This research has been seen to play an important role in increasing the ICT skills of today's individuals and increasing the competencies of ICT teachers in teaching these skills. In the research, it was tried to improve the programming and coding teaching competencies of teachers, which was seen as an important deficiency in the field. With the experimental study conducted after this study, important findings were obtained on how to teach CT skills more effectively. With the goal-based scenario learning method applied in the study, learners comprehended where and under what conditions they could use the knowledge and skills they learned in their real lives. In this method, the content was not presented to the learners as ready-made, but within the scope of the roles and tasks defined to them, it also caused them to use and develop their high-level thinking skills in the process.

In line with the results obtained in this study, some suggestions can be made to researchers and practitioners working in this field:

- The scope of the research is limited to the problem-solving and programming unit of the 6th-grade information technologies course and can be expanded with studies to be conducted at high school and undergraduate levels for the information technologies course.
- The study could be applied in a limited sample due to its experimental structure. Determining the level of CT skills can be applied on larger scale samples.
- More in-depth qualitative studies can be conducted to determine the reasons for the differences identified in the study.
- GBSL can be used in learning environments to improve learners' thinking processes and thinking skills.

References

- Aharoni, D. (2000, March 7-March 12). Cognitive processes of students dealing with data structures. In: *Proceedings of the thirty-first SIGCSE Technical Symposium on Computer Science Education*, Austin, Texas, USA. <https://doi.org/10.1145/331795.331804>
- Aho, A. (2012) Computation and computational thinking. *The Computer Journal*, 55(7),832-835. <https://doi.org/10.1093/comjnl/bxs074>
- Alp, Y. (2019). *The effects of block-based programming on the problem solving ability and attitude towards computer for secondary school students*. (Thesis No. 556354) [Master thesis, İnönü University, Malatya]. Council of Higher Education Thesis Center
- Arabacıoğlu, T., Bülbül, H., & Filiz, A. (2007, January 31-December 2). A new approach to computer programming teaching [Day 1- Session IV]. 9. *Conference of Academic Computing*. Dumlupınar University, Kütahya, Türkiye.
- Arabacıoğlu, T. (2012). *The effect of scenario based teaching programme through different communication media on the achievement of students in information technology class*.

- (Thesis No. 319818) [Doctorate thesis. Aydın Adnan Menderes University-Aydın]. Council of Higher Education Thesis Center
- Atiker, B. (2019). *Effects of computational thinking skills to the success of secondary school students in programming instruction* (Thesis No. 561543) [Doctorate thesis. İstanbul University, -İstanbul]. Council of Higher Education Thesis Center
- Balcı, A. (1995). *Sosyal bilimlerde araştırma: yöntem, teknik ve ilkeler*. (11. Press) Pegem Akademi.
- Bell, B. (1994, April 4-8). The effects of task, database, and guidance on interaction in a goal-based scenario. *American Educational Research Association, New Orleans, LA*. Evanston, IL: Northwestern University
- Bell, B., Bareiss, R., & Beckwith, R. (1993, April 12-16). The role of anchored instruction in the design of a hypermedia science museum exhibit. *Conference of the American Education Research Association, Atlanta, GA*.
- Joint Research Center [JRC] (2016). *Developing computational thinking in compulsory education-implications for policy and practice*. Joint Research Center (European Commission).
http://publications.jrc.ec.europa.eu/repository/bitstream/JRC104188/jrc104188_computhinkreport.pdf.
- Bond, M., & Bedenlier, S. (2019). Facilitating student engagement through educational technology: Towards a conceptual framework. *Journal of Interactive Media in Education, 2019*(1). <https://doi.org/10.5334/jime.528>
- Bransford, J. D., Sherwood, R. D., Hasselbring, T. S., Kinzer, C.K., & Williams, S. M. (1990). Anchored instruction: Why we need it and how technology can help. R. Spiro & D. Nix (Eds.), *Cognition, Education, and Multimedia: Exploring Ideas in High Technology* (pp.115-141). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. Computer Science Teachers Association Task Force.
- Brennan, K., & Resnick, M. (2012, April 13-17). New frameworks for studying and assessing the development of computational thinking. *In Annual Meeting American Educational Research Association*. British Columbia: Vancouver.
- Berry, M. (2013). *Computing in the national curriculum: A guide for primary teachers*. British Computer Society. Retrieved June 11, 2022, from <https://pure.roehampton.ac.uk/portal/en/publications/computing-in-the-national-curriculum-a-guide-for-primary-teachers>
- Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1).32-42. <https://doi.org/10.3102/0013189X018001032>
- Calao, L. A., Moreno-Léon, J., Correa, H. E., & Robles, G. (2015, September 15-18). Developing Mathematical Thinking with Scratch. The 10th *European Conference on Technology Enhanced Learning*, Toledo, Spain. https://doi.org/10.1007/978-3-319-24258-3_2
- Cevahir, H., & Özdemir, M. (2017, May 24-26). Teachers' opinions on the difficulties encountered in programming teaching and suggestions for solutions. *XI. International Computer&Instructional Technologies Syposium*. Malatya, Türkiye

- Ceylan, V. K., & Akar Vural, R. (2023). Self-assessment of computational thinking skill: A scale development study. *Rumeli Journal of Education Studies* (4), 1-22. <https://doi.org/10.5281/zenodo.8217317>
- Ciraj, A.M., Vinod, P., & Ramnarayan, K. (2010). Enhancing active learning in microbiology through case-based learning: experiences from an Indian medical school. *Indian Journal of Pathology and Microbiology*, 53(4), 729-733. <https://doi.org/10.4103/0377-4929.72058>
- Colley, A., & Comber, C. (2003). Age and gender differences in computer use and attitudes among secondary school students: What has changed? *Educational Research*, 45(2), 155–166. <https://doi.org/10.1080/0013188032000103235>
- Cszmadia, A., Curzon, P., Dorling, M., Humphreys, S., Ng, T., Selby, C., & Woollard, J. (2015). *Computational thinking. A guide for teachers*. Computing at School. Charlotte BCS. The Chartered Institute for IT.
- Çatlak, Ş., Tekdal, M., & Baz, F. Ç. (2015). The status of teaching programming with scratch: a document review work. *Journal of Instructional Technologies & Teacher Education*, 4(3).
- Çetin, İ., & Toluk Uçar, Z. (2017). Bilgi işlemsel düşünme tanımı ve kapsamı. Y. Gülbahar. (Ed.). *Bilgi İşlemsel Düşünmeden Programlamaya*. 341-356. Pegem Akademi. <https://doi.org/10.14527/9786052411117>
- Dinçer, A. (2018). *The comparison of 6th grade students' in terms of attitudes ,self-efficacy and academic achievement on teaching of scratch and kodu game lab 'programming languages*. (Thesis No. 512965) [Master Thesis. Dokuz Eylül University-İzmir]. Council of Higher Education Thesis Center
- Errington, R. (2010). What's in a relationship? Exploring cultural assumptions from an international perspective. E.P. Errington (Ed.), *Preparing graduates for the professions using scenario-based learning*. Post Pressed.
- Foster, D., & Bariess, R.A. (1995, April 18-22). Administering the business school case method with a goal-based Scenario. *Annual meeting of the American Educational Research Association*, San Francisco, CA.
- The International Association for the Evaluation of Educational Achievement (IEA). (2014). Preparing for life in a digital age: *The IEA International Computer and Information Literacy Study International Report*. <https://link.springer.com/book/10.1007/978-3-319-14222-7>.
- Gossman, P., Stewart, T., Jaspers, M., & Chapman, B. (2007). Integrating web-delivered problem-based learning scenarios into the curriculum. *Active Learning in Higher Education*, 8(2), pp. 139-153. <https://doi.org/10.1177/1469787407077986>
- Green, J., Caracelli, V., J., & Graham, W.F. (1989). Toward a conceptual framework mixed-method evaluation design. *Educational Evaluation and Policy Analysis*, 11(3). <https://doi.org/10.2307/1163620>
- Guzdial, M., Kay A. Norris, C., & Soloway, E. (2019). Computational thinking should be good thinking. *Communications of the ACM*, 62(1), 28-30. <https://doi.org/10.1145/3363181>
- Gültekin, K. (2006). *The effects of multimedia on computer programming achievement*. (Thesis No: 182319) [Master Thesis, Hacettepe University-Ankara]. Council of Higher Education Thesis Center

- Hoppe, U., & Manske, S. (2022). Developing computational thinking skills with multiple models and representations. S. Kong , H. Abelson & W. Kwok (Eds.), *Computational thinking education in K-12*. MIT Press.
- Hsu, H. J. (2014, February 27-28). Gender differences in Scratch game design. The 3rd *International Conference on Information, Business and Education Technology (ICIBET)*, Beijing, China.
- Hursen, C., & Fasli, F. G. (2017). Investigating the efficiency of scenario-based learning and reflective learning approaches in teacher education. *European Journal of Contemporary Education*. 6(2), 264-279.
- ISTE, (2016). *Computational thinking (Learner). Computational Thinking Competencies*. International Society for Technology in Education (ISTE). <https://iste.org/standards/computational-thinking>
- Kaasbøll, J.J. (1998). *Exploring didactic models for programming*. Proceedings of NIK'98, (Norwegian Computer Science Conference). pp. 195-203
- Kalelioğlu, F., & Gülbahar, Y. (2015, September 9 – 11). What is computational thinking and how to teach it?. *3rd International Instructional Technology and Teacher Education Symposium*. Trabzon, Türkiye,
- Kalelioğlu, F., Gülbahar, Y., Akçay, S., & Doğan, D. (2014, September 22-25). Curriculum integration ideas for improving the computational thinking skills of learners through programming via Scratch. *The 7th International Conference on Informatics in Schools: Situation, Evolution and Perspectives*. Istanbul, Turkey.
- Karabak, D., & Güneş, A. (2013). Curriculum proposal for first class secondary school students in the field of software development. *Journal of Research in Education and Teaching*, 2(3), pp. 173-181.
- Karasar, N. (2013). *Bilimsel Araştırma Yöntemi*. Nobel Yayın.
- Karcı, M. (2018). *Examining the effect of using scenario based teaching method based on STEM activities on students' achievement, career choice and their motivation* (Thesis No. 509021) [Master Thesis, Çukurova University-Adana]. Council of Higher Education Thesis Center
- Kelleher, C., Pausch, R., & Kiesler, S. (2007). Storytelling Alice motivates middle school girls to learn computer programming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, USA, 1455-1464. <https://doi.org/10.1145/1240624.1240844>
- Kemiksiz, C. (2016). *Effects of using scenario-based learning method in 6th grade science classes on academic achievement, attitudes and permanence*. (Thesis No. 429924) [Master thesis, Abant İzzet Baysal University-Bolu]. Council of Higher Education Thesis Center
- Kukul, V. (2018). *The effect of different structured processes on students' computational thinking skills, self-efficacy levels and programming achievements in programming teaching* (Thesis No.527581) [Doctorate thesis, Gazi University-Ankara]. Council of Higher Education Thesis Center

- Malan, D. J., & Leitner, H. H. (2007). Scratch for budding computer scientists. *ACM SIGCSE Bulletin*, 39(1): 223-227.
- Olabe, J. C., Olabe, M. A., Basogain, X., Maiz, I., & Castaño, C. (2011). Programming and robotics with scratch in primary education. A. Méndez-Vilas (Ed.) *Education in a technological world: communicating current and emerging research and technological efforts* in (pp. 356–363). Formatex Research Center.
- Orton, K., Weintrop, D., Beheshti, E., Horn, M., Jona, K., & Wilensky, U. (2016, June 20-24). Bringing computational thinking into high school mathematics and science classrooms. *The International Conference of The Learning Sciences (ICLS): Transforming Learning, Empowering Learners*, Singapore.
- Özmen, B., & Altun, A. (2014). Undergraduate students' experiences in programming difficulties and obstacles. *Turkish Online Journal of Qualitative Inquiry*, 5(3). <https://doi.org/10.17569/tojqi.20328>
- Özvegeç, C., & Kocadağ, H. (2013). The effects of scenario based learning approach to overcome the students' misconceptions about inheritance. *Hacettepe University Journal of Education* 28(3), 83-96.
- Resnick, M., Maloney, J., Hernández, A. M., Rusk, N., Eastmond, E., Brennan, K., Millner, A. D., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y.B. (2009). Scratch: Programming for everyone. *Communications of the ACM*, 52(11), 60–67.
- Saeli, M., Perrenet, J., Jochems, W. M. G., & Zwaneveld, B. (2011). Teaching programming in secondary school: A pedagogical content knowledge perspective. *Informatics in Education*, 10(1), 73-88.
- Samsa, S., Akyüz, H. G., Keser, H., & Numanoğlu, G. (2009, November 17-19). The effect of scenario-based blended learning environments on information technologies and pre-service teachers' attitudes towards teaching profession. 3rd *International Computer and Instructional Technologies Symposium*, Trabzon.
- Sivasakthi, M. & Rajendran, R. (2011). Learning difficulties of object-oriented programming paradigm using Java: students' perspective. *Indian Journal of Science and Technology*, 8(4). 983-985.
- Stolee, K. T., & Fristoe, T. (2011). Expressing computer science concepts through the Kodu game lab. *Association for Computing Machinery, New York, USA*, 99–104. <https://doi.org/10.1145/1953163.1953197>.
- Swanson, H. (2017, July 13-15). Computational thinking in the science classroom. First *International Conference on Computational Thinking Education*. Hong Kong.
- Papert, S. (1980). *Computers For Children. Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc. Publishers.
- Schank, R.C. (1994). Goal-based scenarios: A radical look at education. *The Journal of the Learning Sciences*, 3, 429-453.
- Schoenfeld-Tacher, R., Persichitte, K.A., & Jones, L.L. (2001). Relation of student characteristics to learning of basic biochemistry concepts from a multimedia goal-based scenario. *Journal of Science Education and Technology*, 10(4). 305-317.

- Selby, C., & Woollard, J. (2014, March 5-8). Computational thinking: the developing definition. *The First Conference of Special Interest Group on Computer Science Education (SIGCSE)*, Atlanta, GA.
- Top, O., & Arabacıoğlu, T. (2024). Integrating computational thinking into mathematics education: its effects on achievement, motivation, and learning strategies. *Bayburt Eğitim Fakültesi Dergisi*, 19(42), 2034-2066. <https://doi.org/10.35675/befdergi.1385749>
- Voskoglou, M., G. & Buckley, S. (2012). Problem-solving and computers in a learning environment. *Egyptian Computer Science Journal*, 36(4), 28-46. <https://doi.org/10.48550/arXiv.1212.0750>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49, 33-36.
- Yaman, B. (2005). The effect of drama method in education based on scenario-based learning approach on reading comprehension achievement of fifth grade primary school students. *Journal of Cukurova University Institute of Social Sciences*, 14(2), 465-482.
- Yıldız Durak, H. (2018). Digital story design activities used for teaching programming effect on learning of programming concepts, programming self-efficacy, and participation and analysis of student experiences. *Journal of Computer Assisted Learning* 34(6), 740-752. <https://doi.org/10.1111/jcal.12281>
- Yünkül, E., Durak, G., Çankaya, S., & Mısırlı, Z. (2017). The effects of scratch software on students' computational thinking skills. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, 11(2), 502-517. <https://doi.org/10.17522/balikesirnef.373424>
- Zumbach, J., & Reimann, P. (2002) Enhancing learning from hypertext by inducing a goal orientation: comparing different approaches. *Instructional Science* 30, 243-267. <https://doi.org/10.1023/A:1016009926452>

Genişletilmiş Özet

Günümüzde iş gücü olarak bir takım yenilikçi beceriler aranan özellikler arasında gelmektedir. Bu becerilerin başında da saf bilgi yükü edinmeden ziyade bilgiyi otonom olarak elde edebilen, işleyebilen, farklı düşünme yetilerini içinde barındıran beceriler gelmektedir. Bilgi iletişim dünyasındaki gelişmelerle beraber bilginin insanlığın varoluşundan bu yana hiç görülmediği kadar önemli derecede arttığı (büyük veri) bu dönemde, bilginin üst düzey düşünme becerileri ile işlenmesi daha önemli bir değer olarak kabul edilmektedir. Bu düşünme becerileri arasında gösterilen Bilgi İşlemsel Düşünme (BİD, Computational Thinking), içerisinde barındırdığı çekirdek becerileri ile günümüzde en fazla bahsedilen düşünme alanı olarak karşımıza çıkmaktadır (Brennan ve Resnick, 2012; Berry, 2013). Bu düşünme becerilerini kapsayan öğretim programı MEB 6. Sınıf Bilişim Teknolojileri dersinde 5. ve 6. Sınıflarda “Problem çözme, programlama ve özgün ürün geliştirme” isimli ünite ile eklenmiştir. Bu becerilerin eğitim ortamlarında etkili bir şekilde öğrenenlere kazandırılması noktasında Hedef Temelli Senaryo Öğrenme (HTSÖ) yönteminden yararlanılmıştır. HTSÖ; öğrenenlere açıkça ifade edilmiş bir amaca ulaşmak için problem çözme rolü ve görevini tanımlayan, onlara yaparak öğrenme ortamı sunan bir

yöntemdir. Böylelikle bilgi işlemsel düşünme gibi soyut becerilerin öğrenenlere hayatın içinden gerçek, somut yollarla aktarılması öğretim kalitesini ve verimliliğini arttıracaktır.

Çalışmanın amacı, senaryo temelli Scratch öğretim programının 6. sınıf bilişim teknolojileri dersi problem çözme ve programlama ünitesindeki öğrencilerin akademik başarı puanlarına, öğrenmede kalıcılığına ve BİD becerilerine etkisini değerlendirmektir. Bu temel amaç doğrultusunda aşağıdaki araştırma sorularına yanıt aranmıştır;

1. Deney kontrol grubu öğrencilerinin Bilgi İşlemsel Düşünme Öz Değerlendirme Ölçeği (BİDÖDÖ) ve alt boyutlarının (algoritmik düşünme, paralelleştirme, ayrıştırma, soyutlama ve otomatikleştirme) ön test, son test ve kalıcılık testi toplam puanları karşılaştırıldığında grupların kendi içinde ve gruplar arasında deney grubu lehine test puanları arasında anlamlı farklılık var mıdır ?
2. Deney ve kontrol grubu öğrencilerinin akademik başarı testi ön test puanları kontrol altına alındığında, son test ve kalıcılık testi puanları arasında deney grubu lehine anlamlı farklılık var mıdır?
3. Deney ve kontrol grubu öğrencilerinin akademik başarı ön test puanları kontrol altına alındığında, son test ve kalıcılık testi puanları arasında;
 - o cinsiyet,
 - o ekran kullanım sıklığı
 - o sosyoekonomik durum

açısından deney grubu lehine anlamlı farklılık var mıdır?

Araştırmada nicel yöntemlerden, 2X3 (deney, kontrol grubu X ön test, son test ve kalıcılık testi) yarı-deneyssel modeli kullanılmıştır. Araştırmaya 2018-2019 eğitim öğretim yılında Muğla ili Milas ilçesindeki farklı sosyoekonomik düzeylerden üç okulun 6. sınıfları dahil edilmiştir. Tabaka örnekleme yöntemiyle seçilen her bir okulda bulunan iki adet 6.sınıftan birisi kontrol ve diğer sınıf deney grubu olarak toplam 122 öğrenci belirlenmiştir. Veri toplama aracı olarak 20 maddeden ve 5 boyuttan oluşan Bilgi İşlemsel Öz Değerlendirme Ölçeği (BİDÖDÖ) ve 25 soruluk kapalı uçlu sorulardan oluşan akademik başarı testi kullanılmıştır. Veri toplama araçları uygulama öncesi, uygulama bitiminde ve uygulamada 6 hafta sonra olacak şekilde 3 kez uygulanmıştır. Veri analizi aşamasında hem parametrik hem de parametrik olmayan testler beraber kullanılmıştır. Bu bağlamda akademik başarı testi puanlarının farklı değişkenlere göre farklılık göstermesinin belirlenmesinde tekrarlı ölçümlerde MANCOVA testi kullanılmıştır. BİDÖDÖ ölçeğinin puanlarının gruplar arasında ve grup içinde farklılaşp farklılaşmadığının kontrolünde ise Friedman testi, kontrast analizi ve Wilcoxon Z testi kullanılmıştır. Uygulama Milas ilçesindeki sosyoekonomik düzeyi farklı üç ayrı ortaokulda 13 haftada yürütülmüştür. Bu sürecin üç haftası ön test, son test ve kalıcılık testi uygulamaları için kullanılırken, iki haftası pilot çalışma ve sekiz haftalık süresi ise deneysel uygulama için kullanılmıştır. Bilişim

teknolojileri dersi hem deney hem de kontrol grubunda haftada iki saat olup, her okul türü için haftalık dört saat kontroller yapılmıştır. Uygulamayı hem deney hem de kontrol gruplarında okullarda görevli Bilişim Teknolojileri öğretmenleri araştırmacının rehberliğinde yürütmüştür. Çalışmada deney grubu öğrencilerine senaryo temelli Scratch öğretim programı uygulanırken kontrol grubuna uygulayıcı öğretmenlerin daha önceki süreçte kullandıkları mevcut MEB Programı uygulanmıştır. MEB programı kapsamında öğretmenler ünitenin işlenişinde fişsiz kodlama etkinlikleri ve Scratch aracını kullanmışlardır. Uygulanan bu programda ki en temel fark öğretmenlerin geliştirdikleri Scratch programlarını öğrencilere doğrudan sunmaları ve gösterip yaptırma ile uygulamaları olmuştur. Öğrencilerden bu programları kendi bilgisayarlarında bazen bireysel bazen grup ile geliştirmeleri istenmiştir. Soru cevap, gösterip yaptırma, anlatım tekniklerinin ağırlıklı olarak kullanıldığı bir uygulama olmuştur. Senaryo temelli Scratch öğretim programının geliştirme sürecine araştırmacının yanı sıra iki öğretim üyesi, üç bilişim teknolojileri öğretmeni ile bir program geliştirme uzmanı aktif olarak katılım sağlamıştır. Öğretim programı kapsamında toplam sekiz adet senaryo geliştirilmiştir.

Araştırma bulgularında; HTSÖ ile öğrenim gören deney grubu öğrencileri ile kontrol grubu öğrencilerinin akademik başarı son test ve kalıcılık testi puanlarında deney grubu lehine anlamlı fark bulunmuştur. Çalışmada bu durum farklı değişkenler içinde kontrol edilmiştir. Bu değişkenler sosyoekonomik okul düzeyi, cinsiyet ve ekran kullanım sıklığıdır. Bulguların bu değişkenlere göre farklılaştığı görülmüştür.

Deney ve kontrol grubu arasında ortaya çıkan farkı incelemek amacıyla son test, kalıcılık testi puanları ve grupların ortalama puanları kontrol edilmiştir. Buna göre deney grubundaki öğrencilerin akademik başarı son test ve kalıcılık testi düzeltilmiş ortalama puanlarının kontrol grubundaki akademik başarı son test ve kalıcılık testi düzeltilmiş ortalama puanlarından daha yüksek ve anlamlı farklılığa sahip olduğu tespit edilmiştir. Akademik başarı testinin tekrarlı ölçümlerinde ekran kullanım sıklığı haftalık 0-1 saat olanların 3 saat ve üzeri olanlara göre anlamlı olduğu ve test puanlarının daha yüksek olduğu belirlenmiştir. Akademik başarı testi puanları gruplar arasında farklı üst düzey sosyoekonomik okuldaki deney grubu ile alt düzey sosyoekonomik okuldaki deney ve kontrol grupları arasında üst düzey sosyoekonomik deney grubu lehine anlamlı farklılık taşımaktadır. Kontrol grubu öğrencilerinin BİDÖDÖ son test ve kalıcılık testi ortalama puanlarına göre anlamlı farklılık bulunmazken, test puanları arasında deney grubunun puanlarının farklılaştığı ve yalnızca soyutlama alt ölçeği puanlarına göre gruplar arasında deney grubu lehine anlamlı farkın olduğu tespit edilmiştir.