Journal name	International e-Journal of Educational Studies
Abbreviation	IEJES
e-ISSN	2602-4241
Founded	2017
Article link	http://doi.org/10.31458/iejes.1240193
Article type	Research Article
Received date	21.01.2023
Accepted date	25.05.2023
Publication date	31.07.2023
Volume	7
Issue	14
pp-pp	287-309
Section Editor	Prof.Dr. Kürşat YENİLMEZ
Chief-in-Editor	Prof.Dr. Tamer KUTLUCA
Abstracting & Indexing	Education Source Ultimate Database Coverage List EBSCO Education Full Text Database Coverage List H.W. Wilson Index Copernicus DRJI Harvard Library WorldCat SOBIAD
Article Name	Collaborative Testing as an Alternative Assessment Technique in Algebra Education

Author Contribution Statement

[1] **Selcen ÇALIK UZUN** Assist.Prof.Dr. *Trabzon University, Turkey*	Conceptualization, literature review, methodology, implementation, data analysis, translation, and writing
[2] **Sedef ÇELİK DEMİRCİ** Assist.Prof.Dr. *Artvin Çoruh University, Turkey*	Conceptualization, literature review, methodology, implementation, data analysis, translation, and writing

Abstract

This study aims to analyze the impact of collaborative testing of students' cognitive and affective learning outcomes in learning algebra in mathematics education. The sample of the study was 33 students of 6th grade. The mixed method approach was employed, using both quantitative and qualitative data collection techniques. Data collection instruments included (a) exams consisting of open-ended questions developed by the researchers to test student attainments in algebra (Algebra Attainment Test), (b) Mathematical Attitudes Scale aiming to determine student attitudes towards math lessons, (c) Exam Anxiety Inventory aiming to determine student exam anxiety levels, and (d) interviews aiming to determine students' opinions on collaborative testing. According to the analysis of the quantitative data, it was determined that the experimental group's mean score on the algebra attainment test was higher than the control group. But it was not statistically significant. Furthermore in the total mean score of the exam anxiety was found a decrease in the experimental group but it was not significant. However, it was determined that there was a significant difference between the attitudes of the students in the experimental group towards mathematics. The analysis of the qualitative study data revealed that students' views were both positive and negative on the "cognitive", "affective", "social" and "suggestions" themes about collaborative testing. As a result of the research, we evaluated the effectiveness of collaborative testing in the form of group work based on assessment activities. At the same time, we presented the advantages and disadvantages of the technique and discussed its usability as an alternative assessment technique.

To cite this article:

Çalık-Uzun S., & Çelik-Demirci, S. (2023). Collaborative testing as an alternative assessment technique in algebra education. *International e-Journal of Educational Studies, 7* (14), 287-309. https://doi.org/10.31458/iejes.1240193

Research Article

Collaborative Testing as an Alternative Assessment Technique in Algebra Education [*]

Selcen ÇALIK UZUN [1] **Sedef ÇELİK DEMİRCİ** [2]

Abstract

This study aims to analyze the impact of collaborative testing of students' cognitive and affective learning outcomes in learning algebra in mathematics education. The sample of the study was 33 students of 6th grade. The mixed method approach was employed, using both quantitative and qualitative data collection techniques. Data collection instruments included (a) exams consisting of open-ended questions developed by the researchers to test student attainments in algebra (Algebra Attainment Test), (b) Mathematical Attitudes Scale aiming to determine student attitudes towards math lessons, (c) Exam Anxiety Inventory aiming to determine student exam anxiety levels, and (d) interviews aiming to determine students' opinions on collaborative testing. According to the analysis of the quantitative data, it was determined that the experimental group's mean score on the algebra attainment test was higher than the control group. But it was not statistically significant. Furthermore in the total mean score of the exam anxiety was found a decrease in the experimental group but it was not significant. However, it was determined that there was a significant difference between the attitudes of the students in the experimental group towards mathematics. The analysis of the qualitative study data revealed that students' views were both positive and negative on the "cognitive", "affective", "social" and "suggestions" themes about collaborative testing. As a result of the research, we evaluated the effectiveness of collaborative testing in the form of group work based on assessment activities. At the same time, we presented the advantages and disadvantages of the technique and discussed its usability as an alternative assessment technique.

Keywords: Alternative measurement and assessment, collaborative testing, algebra learning domain, cognitive learning affective learning

1. INTRODUCTION

Measurement and assessment is a significant element in curricula. It is known that in traditional curricula, measurement and assessment activities aim to measure knowledge-based attainment of students with exam questions (Kuran & Kanatlı, 2010). In traditional measurement and assessment, answers to a series of questions are evaluated within a certain timeframe, and learning experiences are neglected (Anderson, 1998). The reason is that traditional methods assess student attainment as separate from the instruction process and prioritize outcomes (Gelbal & Kelecioğlu, 2007). In other words, as traditional testing focuses on the assessment of final learning outcomes, an evaluation of learning experiences are often ignored (Baki & Birgin, 2004; Çoruhlu, Nas & Çepni, 2009). Cansız-Aktaş (2018) underlines that during the assessment phase of learning the student's effort, that is, the process, should be taken into consideration in the production of this product, as well as the product produced by the student. Developments in epistemol ogical theories introduced new measurement and assessment approaches in learning (Baki & Birgin, 2002). Alternative measurement and assessment

Received Date: 21/01/2023 **Accepted Date:** 25/05/2023 **Publication Date:** 31/07/2023

[*] ***To cite this article:*** Çalık-Uzun S., & Çelik-Demirci, S. (2023). Collaborative testing as an alternative assessment technique in algebra education. *International e-Journal of Educational Studies, 7* (14), 287-309. https://doi.org/10.31458/iejes.1240193.

[1] *Assist.Prof.Dr.,Trabzon University, selcencalikuzun@trabzon.edu.tr, Trabzon, Türkiye*

[2] *Assist.Prof.Dr.,Artvin Çoruh University, sedefcelik@artvin.edu.tr, Artvin, Türkiye*

Corresponding Author e-mail adress: selcencalikuzun@trabzon.edu.tr

approaches contribute to student learning through feedback mechanisms on the progress and challenges experienced by students, in addition to grading (Karamustafaoğlu, Çağlak, & Meşeci, 2012). Thus, to be able to determine any challenges to learning, assessment methods such as experiments, presentations, exhibitions, projects, discussions, observation, interviews, portfolios, self-assessment and peer assessment, should also be conducted in addition to written tests/exams (Toptaş, 2011). The fact that alternative measurement and assessment approaches allow the analysis of skills and attitudes, as well as knowledge (Çalışkan & Kaşıkçı, 2010) increasing its significance (Baki & Birgin, 2004; Duban & Küçükyılmaz, 2008; Gelbal & Kelecioğlu, 2007).

In recent years, collaborative testing came to the fore as one of the alternative measurement and assessment approaches. In learning activities conducted as groupwork, students are actively involved in the process of construction of knowledge (Quarstein & Peterson 2001). In the group exam technique, it is possible to ensure that students are involved in assessment activities in the form of group work, and to allow them to work together in the process of answering the questions. By doing so, in addition to assessing learning outcomes resulting from the activities they complete together with their peers, it will also be possible to assess throughout the learning process. More detailed information about Collaborative Testing is given below.

1.1. Collaborative Testing

Collaborative testing entails solving the exam questions with collaboration between the students (Lusk & Conklin, 2003). Collaborative testing was also called the two-stage testing (Hendricson, Brady, & Algozzine, 1987) or pyramid testing (Yuretich, Khan, Leckie, & Clement, 2001) in different researches. There are a variety of approaches in regard to the practice of collaborative testing to be found in the literature. The most common approach was for students to take individual exams after the collaborative test, or vice versa (Breedlove, Burkett, & Winfield, 2007; Lusk & Conklin, 2003; Zipp, 2007). Another method proposes to allow students to take the same test twice, first individually and then as a group (Giuliodori, Lujan, & DiCarlo 2008; Ioannou & Artino, 2010; Rao, Collins, & DiCarlo, 2002).

A review of the literature on collaborative testing found that studies revealed the impact of this method on student learning (Breedlove et al., 2007; Bloom, 2009; Guiliodori et al., 2008; Muir & Tracy, 1999; Zimbardo, Butler, & Wolfe, 2003; Zipp, 2007). Researchers (Hodges, 2004; Giuliodori et al., 2008; Zipp, 2007) indicated that collaborative testing was one of the methods that could be employed in the learning and instruction and argued that the technique could be used in the construction of cognitive knowledge, as well as assessment. Zimbardo et al. (2003) claimed that students who were tested collaboratively achieved higher success when compared to individual tests. Thus, the implementation of the collaborative testing in learning areas where students experience difficulties could improve the recognition of the attainments in the learning area. Thus, algebra, as one of the difficult learning topic areas, was selected in the current study. Collaborative testing leads to an improvement in interaction among the students (Giraud, 1997; Ioannou & Artino, 2010; Keller & Steinhorst, 1995; Magel, 1998), allowing the discussion of the difficult-to-understand areas of attainment in algebra. Furthermore, it is suggested that acquisitions induced by the feedback provided in collaborative testing would positively reflect on the individual test/exam performances to be administered later on (Mahoney & Harris-Reeves, 2019).

In addition to the learning outcomes created by students, an important stage is to assess how performance is affected by perception during a group exam and how behaviors affect exam performance. Exam anxiety, which is significant among these behaviors, is described as the emotional state that leads to stress in the individual during assessment activities, preventing the real performance of the individual (Spielberger, 1995). Exam anxiety can cause some negative situations such as fear or anxiety and not being able to experience the exam process as desired (Schutz, Distefano, Benson, &

Davis 2004). Exam anxiety tends to weaken students' abilities to successfully sit exams and, ultimately, their overall grades (Cantwell, Sousou, Jadotte, Pierce, & Akioyamen, 2017). In Türkiye, many students develop negative attitudes towards mathematics lessons based on the idea that mathematics is difficult after the primary education, which leads to exam anxiety in mathematics. This situation leads to exam anxiety in the context of mathematics lessons (Dursun & Bindak, 2011; Yenilmez, Girginer, & Uzun, 2007). Several studies reported on the positive impact of collaborative testing on exam anxiety. Breedlove et al. (2007) reported positive outcomes of collaborative testing, which improved student achievements by reducing student stress and anxiety. Collaborative testing allows students to share or affirm their answers by other students in the group, thereby reducing anxiety and facilitate the recall of knowledge, thus leading to higher grades (Mitchell & Melton, 2003). Other studies reported that students participating in collaborative tests exhibited lower anxiety levels during learning in the classroom setting and also exams (Lusk & Conklin, 2003). In fact, it is known that students experience more exam anxiety in challenging learning areas, such as algebra (Reyes & Castillo, 2015). An examination of students' attitudes towards learning algebra should also be undertaken. In fact, Çalık-Uzun and Birişçi (2018) found that students' motivation to participate in class activities increased with the collaborative testing technique in mathematics. In this study, it is thought that the use of collaborative testing in the field of algebra learning will contribute to the field area.

It is known that other types of attitudes towards subjects that develop during learning activities in lessons also develop. Findings reported by studies conducted on student attitudes reported that students developed positive attitudes after collaborative testing processes (Giraud & Enders, 2000; Ioannou & Artino, 2010). Further, Slusser and Erickson (2006) reported that collaborative testing practices affected student attitudes towards the subject/lesson, increasing motivation.

1.2. The Purpose of the Study

The relevant literature review revealed several studies on the effectiveness of the cooperative learning strategies. However, in Türkiye the number of studies on collaborative testing (which is frequently employed in cooperative learning) is limited, and there are only a few studies on collaborative testing practices in the construction and evaluation of knowledge. Çalık-Uzun and Birişçi (2018) investigated the teacher and student views on collaborative measurement and reported that teachers and students had positive views on collaborative testing; it was found that collaborative testing increases their motivation for participation in the lesson. In this study, in addition to teacher and student opinions, it is aimed to examine the cognitive and affective learning outcomes with the collaborative testing technique. When the studies conducted regarding the collaborative testing technique in the literature are examined, it is seen that the collaborative testing technique is mostly discussed in terms of its contribution to learning (Lusk & Conklin, 2003; Rao et al., 2002). Therefore, it is thought that this study, in which the effect of the collaborative testing technique on cognitive and affective learning outcomes will be investigated, is more comprehensive. On the other hand, a study that examined the collaborative testing technique in terms of exam anxiety, interaction between groups, etc. was conducted in the statistics lesson (Kapitanoff & Pandey, 2018). However, this study will be conducted on learning outcomes in algebra in lower secondary education. Therefore, it is thought that this study that implements collaborative testing will provide a different perspective to the evaluation of the objectives in the field of algebra learning, which is quite abstract for lower secondary students. It was suggested that the implementation of both individual and collaborative tests would be an alternative to both formative and complementary activities. Thus, collaborative testing could be an alternative to measurement and assessment activities. It was suggested that this technique, which is quite different from traditional assessment techniques, would lead to cognitive and permanent learning, and ensure active student participation. Furthermore, the interaction between the group student members and the student during the test would contribute to student attitudes towards the subject and the topic and reduce student anxiety. The collaborative testing allows the students to learn within a discussion environment. Thus, collaborative testing as an alternative measurement and assessment method would also provide information on the learning process. Therefore, the main research problem was determined as follows: "*Does collaborative testing affect students' cognitive and affective learning outcomes in algebra?*" Therefore, the study aimed to investigate the impact of the collaborative testing activities on cognitive and affective learning in the 6th grade mathematics subject in algebra lessons. The cognitive learning of the students was limited to their academic performance, and affective learning was limited to student attitudes and anxiety. Thus, responses to the following sub-problems were sought:

1. What is the effect of collaborative testing on students' academic performance in algebra?
2. What is the effect of collaborative testing on affective learning?
    a) Is there any effect of collaborative testing on students' exam anxiety?
    b) Is there any effect of collaborative testing on students' attitudes towards mathematics?
3. What are the students' views on the collaborative testing process?

2. METHOD

2.1. Research Design

The present study aimed to investigate the impact of collaborative testing on the students' cognitive and affective learning outcomes. The mixed method approach was employed to determine the effects of collaborative testing on the academic achievement, attitudes and exam anxiety levels of students. Plano Clark and Creswell (2008) and Johnson and Onwuegbuzie (2004) reported that the mixed method allows various data collection instruments to test in-depth the research questions. The current study adopted the explanatory sequential design, a mixed method, to investigate the students' cognitive and affective learning outcomes. In this design, quantitative data are collected and analyzed as a first step. Then, participant responses in the quantitative dimension are discussed based on the interviews conducted in the qualitative dimension (Creswell, 2016). It is considered that it would be appropriate to use the quantitative method to solve the 1st and 2nd questions of the research. And so, for the first question, the static group comparison design, which is sort of a weak experimental design (Büyüköztürk, Çakmak, Akgün, Karadeniz, & Demirel, 2020), was adopted to reveal the effect of collaborative testing on student achievement in algebra. On the other hand, the one-group pretest-posttest design which is another sort of weak experimental design was used to test the second research question.

In the study, qualitative data were collected to investigate students' views of collaborative testing and complement the quantitative data. Therefore phenomenology design is the sort of qualitative design was used to test the third research question. Phenomenological research is conducted to explore the experiences of several people with a concept (Creswell, 2016). This study aimed that explore students' experiences of collaborative testing process.

Tablo 1. Research design

Mixed Method		
Quantitative	Weak Experimental Design	
Testing First research problem	The Static-Group Comparison Design	
Group	Pre-test	Post-test
Experimental	-	+
Control	-	+
Testing second research problem	The One Group Pretest-Posttest Design	
Group	Pre-test	Post-test

Experimental Group	+	+
Qualitative	Phenomenology Design	
Testing third research problem	Phenomenology Design	
Group	Experimental Group	

2.2. Study Sample

The current study was carried out with 33 students, 16 of which were from 6-A class and 17 of which were from 6-B class and their mathematics teachers in a lower secondary high school in an urban center in the Eastern Black Sea Region in Türkiye. The mathematics performance scores of students in the different classrooms were compared and two equivalent classrooms were selected by also receiving the opinions of the math teachers. The groups were randomly assigned as the experimental and control groups.

2.3. The Research Process

The research process was designed in two stages. The first stage aimed to determine the learning areas associated with testing activities during the teaching process. The collaborative testing activities were conducted in the 6th grade mathematics lessons. After official approval was obtained, face-to-face interviews were conducted with the mathematics teachers in the school. In these interviews, mathematics teachers stated that, based on their prior experiences, students experienced difficulties in acquiring algebra attainments and thus, the activity was designed for algebra. The content of the testing activities was developed to include predetermined learning area acquisitions.

In the second stage involved designing the collaborative testing activities which would be applied throughout the research. In this design process, the number of attainments for algebra was significant. The planned collaborative tests were modelled initially as collaborative, and then, as individual exams. Both collaborative and individual tests were designed to include similar questions that measure the same attainment. Testing focused on student performance in the application of knowledge in similar situations. The collaborative testing model developed in the study is presented in Table 2.

Table 2. The collaborative testing model developed in the study

1st Quiz	Feed back	2nd Quiz	Feed back	3rd Quiz	Feed back	1st exam	4th Quiz	Feed back	5th Quiz	Feed back	2nd exam	Follow-up Test

:Individual test :Collaborative test

The activities were conducted in 8 weeks. Since it was required that the students have subject-area knowledge on related attainments prior to the testing, the time schedule was planned with the mathematics teacher. The researchers did not intervene in the instruction process. An introductory meeting about the study was held with the students in the experimental group. In the prepared presentation, the students' questions about the process were answered sincerely since they would be part of such a study for the first time. The activities were conducted by the researchers during the class hours allocated for this purpose every week. Çalık Uzun and Birişçi (2018) suggest that there should be a feedback process in order to conduct the collaborative testing more effectively. Considering this suggestion, a feedback phase was added to the implementation process while planning this study. After each test, student responses were examined by the researchers; student misconceptions, errors and problems in these questions were identified. Before the next exam, separate interviews were conducted with the groups to address and eliminate these issues. Since the individual test and the group test were prepared as equivalent, it was not necessary to provide feedback after the individual test. The mathematics teacher did not participate in the implementation session. One week after the activities, interviews were conducted with the all students about their experiences during the implementation.

2.4. Data Collection Instruments

In accordance with the nature of the mixed method approach, both quantitative and qualitative data collection techniques were employed in the present study. It was considered that the quantitative data could be further elaborated with the qualitative data. However, the algebra attainment test could not be applied as pre-test in the study group. It would not be adequate to measure the academic achievement in algebra as a pre-test since algebra was included in the 6th grade for the first time and the students did not have any early algebra knowledge. Considering that the participants moved from an early-algebraic period to an algebraic period and encountered the learning outcomes related to this learning area for the first time, it was thought that it would not be meaningful to conduct a pre-test for algebra performance in the experimental group. The holistic version of the quantitative data collection instruments is presented in the Table 3 below.

Table 3. Quantitative data collection instruments and their implementation

Group	Pretest	Implementation	Posttest
Experimental	-	Collaborative Testing	Algebra attainment test
Control	-	-	Algebra attainment test
Experimental	Mathematical attitude test	Collaborative Testing	Mathematical attitude test
	Exam anxiety inventory		Exam anxiety inventory

On the other hand qualitative data were collected by interviewing students that joined the study. In the interviews, the students were asked questions about their experiences in the collaborative testing process. Sample questions can be listed as follows; What are your likes and dislikes about the collaborative test technique? Please describe your experiences; Is there anything you learned from your friends during this implementation process? Can you explain with examples?; You have joined tests individually and with a groupmate for a few weeks What can you say when you compare the individual tests with the collaborative tests? The data collection tools to be used within the sub-problems of the study and their explanations are given in detail below under sub-headings.

2.4.1. Algebra attainment test

The open-ended exams on algebra for both collaborative and individual testing activities were developed by the researchers and the mathematics teacher of the class where the implementation is carried out. First of all, open-ended parallel questions were prepared with the mathematics teacher for algebra achievements. And then 2 experts in mathematics education were consulted about these questions regarding their suitability for the achievements and their equivalents of questions. Then, these questions were implemented to 5 randomly selected 7th grade students. After the implementation the questions the participants failed to comprehend were revised. A similar method was adopted in all tests. Since the final questions which were arranged according to the feedback and prepared in accordance with the attainments were similar, they were randomly assigned to group and individual activities. These exam included 6 attainments in algebra sub-learning areas in the mathematics syllabus. The attainment included in the tests are presented in Table 4.

Table 4. Algebra attainments included in the tests

Test type	Application	Related attainment					
1st Quiz	Collaborative	6.2.1.1.					
2nd Quiz	Collaborative	6.2.1.2.	6.2.1.3.				
3rd Quiz	Collaborative	6.2.1.4.					
1st Exam	Individual	6.2.1.1.	6.2.1.2.	6.2.1.3.	6.2.1.4.		
4th Quiz	Collaborative	6.2.1.5.					
5th Quiz	Collaborative	6.2.1.6.					
2nd Exam	Individual	6.2.1.5.	6.2.1.6.				
Follow-up test	Individual	6.2.1.1	6.2.1.2.	6.2.1.3.	6.2.1.4.	6.2.1.5.	6.2.1.6.

6.2.1.1. Expresses the rule in arithmetic sequences with letters, can determine the requested term in the arithmetic sequence depicted with letters.
6.2.1.2. Can express verbally in algebraic terms and can express an algebraic case verbally.
6.2.1.3. Can calculate an algebraic equality based on various natural number variations.
6.2.1.4. Can discuss the meaning of simple algebraic expressions.
6.2.1.5. Can add and subtract algebraic representations.
6.2.1.6. Can multiply an algebraic expression by a natural number.

Each activity was implemented as collaboratively or individually when the time came during the research process. Examples of parallel questions are given in Figure 2 and Figure 3.

Fill in the blanks in the table below with appropriate expressions.

Algebraic Expression	Coefficient	Constant Term	Number of Term
$\frac{3m}{8}-54$			
	-11	67	
$17z-106$			
	2	-2,2	2
	-0,6	4	
$0,07x-23$			

Fill in the blanks in the table below with appropriate expressions.

Algebraic Expression	Coefficient	Constant Term	Number of Term
$5x$			
	3	2	2
$\frac{y}{4}-5$			
	-6	-3	
$13-5z$			
	0,7		1

Figure 1. Parallel question examples developed for collaborative and individuals tests

Hilal is considering changing her mobile phone contract. After examining the different contracts, she gets the following information about them.

Contract A
15 TL Fixed Fee
&
45 cents per minute

Contract B
60 cents per minute

According to the given information above, assuming that Hilal speaks at an average of x minutes per month, fill in the blanks in the table below with the appropriate algebraic expressions.

Talk time (minute)	Fee	
	Contract A	Contract B
x		

Arzu decides to go to the gym. She receives fee information from 2 gyms in the city. Since Arzu thinks of going to the gym "**x**" times a month, find the algebraic expressions that give the price she has to pay for each gym.

Efor Gym — ***algebraic expressions***
6 Turkish Lira
for each session

Fit Gym — ***algebraic expressions***
20 TL fixed fee
&
2 Turkish Lira
for each session

Figure 2. Parallel question examples developed for collaborative and individuals tests

2.4.2. Exam anxiety inventory

The Exam Anxiety Inventory (EAI), developed by Spielberger in 1980 and adapted to Turkish language by Öner (1990), was used to determine exam anxiety levels of the students. As suggested by Liebert and Morris (1967), the inventory includes two sub-dimensions: "emotionality" and "worry". EAI includes 20 items that are scored based on a 4-point scale [(1) almost never, (2) sometimes, (3) often, (4) almost always] with instructions. In the Turkish language version, the worry dimension includes 8 items (2, 3, 4, 5, 8, 12, 17, 20), and the emotionality dimension includes 12 items (1, 6, 7, 9, 10, 11, 13, 14, 15, 16, 18, 19). The highest score that can be obtained is 80, and the lowest score is 20.

2.4.3.Mathematical attitude scale

The Mathematics Attitude Scale (MAS) scale was developed by Aşkar (1986) to determine student attitudes towards mathematics. The MAS includes 10 positive and 10 negative 5-point scales ("Strongly Disagree", "Disagree", "Undecided", "Agree" and "Strongly Agree"). The Cronbach Alpha reliability coefficient of the scale developed by Aşkar (1986) is 0.96 and it is a single dimensional scale. The Cronbach Alpha reliability coefficient was calculated as 0.93 in the current study.

2.4.4. Interviews

In this study, the interview technique was employed to determine the views of the students on the collaborative testing. Face-to-face interviews that included semi-structured questions were conducted with the participants and interviews were recorded after participant consent was obtained. Although the questions in semi-structured interviews were predetermined, this technique was considered advantageous since the interviewers could change the order of the questions and provide the opportunity to respond to the questions in detail (Çepni, 2009). The semi-structured interview questions were developed by the researchers and asked individually to the students who participated in the collaborative test. To test the relevance and comprehensibility of the questions, they were submitted to two mathematics education experts for their review. The researchers asked these experts to make suggestions about the questions that should be included in the interview questions. For example, suggestion of adding questions about what was good and what were the difficulties that students face when doing group work to the interview questions were taken into consideration. The interview questions were then revised and finalized based on their expert opinions. Interviews lasted approximately 30 minutes. The recorded interviews were transcribed.

2.5. Data Analysis

Quantitative and qualitative data collection techniques were employed in the research methodology, and different analysis methods were used to analyze the data. How the analysis of data collected is explained below.

2.5.1. Quantitative data analysis

SPSS 17.0 program was used in the analysis of the quantitative data, arithmetic mean, standard deviation and percentages were calculated. Normality analyzes were applied to decide on the use of parametric and non-parametric tests. It is known that the Shapiro-Wilks method, which is one of the methods used when evaluating normality, is statistically powerful in small samples (Pituch & Stevens, 2009). In this study, Shapiro-Wilks normality analysis was used since there were 16 in the experimental group and 17 in the control group. In addition, the normal distribution of the data was evaluated by calculating the skewness and kurtosis values. It was seen that the quantitative data for the research's cognitive learning outputs were normally distributed. Normality values are given in Table 5.

Tablo 5. Normal distribution of data for cognitive learning

Shapiro-Wilks Test	N	p	Skewness	Kurtosis
Experimental	16	0.39	-0.48	-0.32
Control	17	0.11	-0.29	-1.24

As can be seen in Table 5, the data for cognitive learning outcomes are normally distributed ($p > 0.05$). It is stated that skewness and kurtosis values between +2 and -2 are seen as an adequate criterion for normality (Gravetter & Wallnau, 2013). In addition, histogram graphs and Q-Q plot graphs related to the normal distribution of the control and experimental group data are included in the appendices of the research (Appendix-A). İndependent samples t-test was used to determine the differences between the control and experimental groups at p=0.05 significance level. On the other hand this research is a static group comparative design, the data on affective learning were examined within the experimental group itself. The normality values of the data for affective learning were examined in order to make analyzes for in-group evaluation. Normality values are given in Table 6.

As can be seen, the data on mathematical attitude and test anxiety, which are among the data for affective learning, are normally distributed. The data for affective learning outcomes are normally distributed ($p > 0.05$).

Tablo 6. Normal distribution of data for affective learning

Shapiro-Wilks Test	N	p	Skewness	Kurtosis
Exam anxiety	16	0.60	0.14	0.96
Mathematical Attitudes	16	0.25	0.98	0.75

In addition, histogram graphs and Q-Q plot graphs related to the normal distribution of the mathematical exam anxiety and attitude data are included in the appendices of the research (Appendix-B-C). Therefore, the Dependent Samples t-test was used to look at the test anxiety and mathematics attitude of the experimental group within itself. When the sub-dimensions of the anxiety scale for exams were examined, it was observed that the data for each dimension were also normally distributed (p_{worry} and $p_{emotionality} > 0.05$).

2.5.2. Qualitative data analysis

The qualitative data collected from interviews from students were analyzed with content analysis (Yıldırım & Şimşek, 2008). Thus, initially, the interview records were transcribed. In doing so, the statements of the students and the teacher were transcribed verbatim as they were expressed during the interview so that they remained structurally intact. In the analysis, the transcribed interview data were transferred to the MAXQDA 2020 qualitative data analysis software and coded separately by both researchers. Researcher triangulation and time triangulation techniques were used to increase validity. Researcher triangulation means including more than one researcher in the data analysis process of the study (Başkale, 2006). Both researchers separately coded at different time intervals, and the codes were brought together and discussed until a consensus was reached. It was then decided under which themes the common codes obtained should be grouped. The determined codes and themes were employed in the analysis, and the findings were interpreted and supported by direct quotes. In the process, code names (S1, S2, etc.) were used instead of the actual names of students.

3. FINDINGS

The findings of this study, which was conducted to determine how the exam activities carried out collaboratively affected learning outcomes, are presented as items in line with the sub-objectives of the research.

3.1. Cognitive Learning Findings

This section outlines the findings on the tests administered in the experimental (6A) and control (6B) groups after the unit lesson was taught. The mean test scores and comparison of the group scores were determined with the independent samples t-test, and the results are presented in Table 7.

Table 7. Independent samples t-test results of algebra exam test scores of students in experimental and control groups

Group	**N**	**M**	**SD**	***df***	***t***	***p***
Experimental (6A)	**16**	**22.5**	**12.24**	**31**	**-5.33**	**0.59**
Control (6B)	**17**	**20.29**	**11.52**			

The analysis of the data presented in Table 7 revealed that the mean algebra attainment test score of the experimental group (*M* = 22.5, *SD* = 12.24 was higher than that of the control group (*M*= 20.29, *SD* = 11.52) in the collaborative test. It was found that the difference between the mean scores in terms of increased performance was not statistically significant ($p > .05$).

3.2. Affective Learning Findings

3.2.1. Exam anxiety scale (EAS) findings

The impact of collaborative testing on exam anxiety was tested according to the second sub-problem. To determine the whether there was a statistical significance between the mean scores, Dependent groups t test was carried out and the findings are presented in Tables 8.

Table 8. Dependent samples t-test results of exam anxiety test scores

Exam Anxiety Test	Test	N	*M*	*SD*	*df*	*t*	*p*
Worry	Pre-test	**16**	**17.87**	**6.31**	**15**	**0.32**	**0.74**
	Post-test	**16**	**18.43**	**6.58**			
Emotionality	Pre-test	**16**	**28.50**	7.27	**15**	**-0.66**	**0.51**
	Post-test	**16**	**27.50**	**6.48**			
Total score (EAS)	Pre-test	**16**	**46.37**	**12.54**	**15**	**-0.15**	**0.88**
	Post-test	**16**	**45.93**	**12.65**			

The analysis of the mean sub-dimension scores presented in Table 8 demonstrated that the anxiety levels in the "worry" sub-dimension increased in the post-test, while the anxiety levels decreased in the "emotionality" dimension. The total mean pre-test score was 46.37 and the total mean post-test score was 45.93. However, the decrease in exam anxiety was not statistically significant in the experimental group ($p > .05$).

3.2.2. Mathematical attitude scale (MAS) findings

The impact of the collaborative testing on student attitudes towards mathematics was tested in regard to the third sub-problem. To determine the statistical significance of the increases in mean scores, dependent groups t test was conducted and the results are presented in Tables 9.

Table 9. Dependent samples t-test results of mathematical attitude test scores

Mathematical Attitude Test	Test	N	*M*	*SD*	*df*	*t*	*p*
Total score (MAS)	Pre-test	**16**	**66.62**	**12.88**	**15**	**-4.49**	**0.00**
	Post-test	**16**	**78.93**	**10.18**			

The mean pre-test scale score was 66.62 and mean post-test scale score was 78.93. The general analysis of the MAS scores revealed that the increase in the post-test attitude scores towards mathematics was statistically significant after collaborative testing ($p < .05$).

3.3. Students' Views Findings

The views of the students on the collaborative testing were categorized in four themes: "cognitive", "affective", "social" and "suggestions" (see Figure 3). It was observed that the views of the participating students in the cognitive dimension of the collaborative testing application were grouped under categories such as "individual learning" and "collaborative learning".

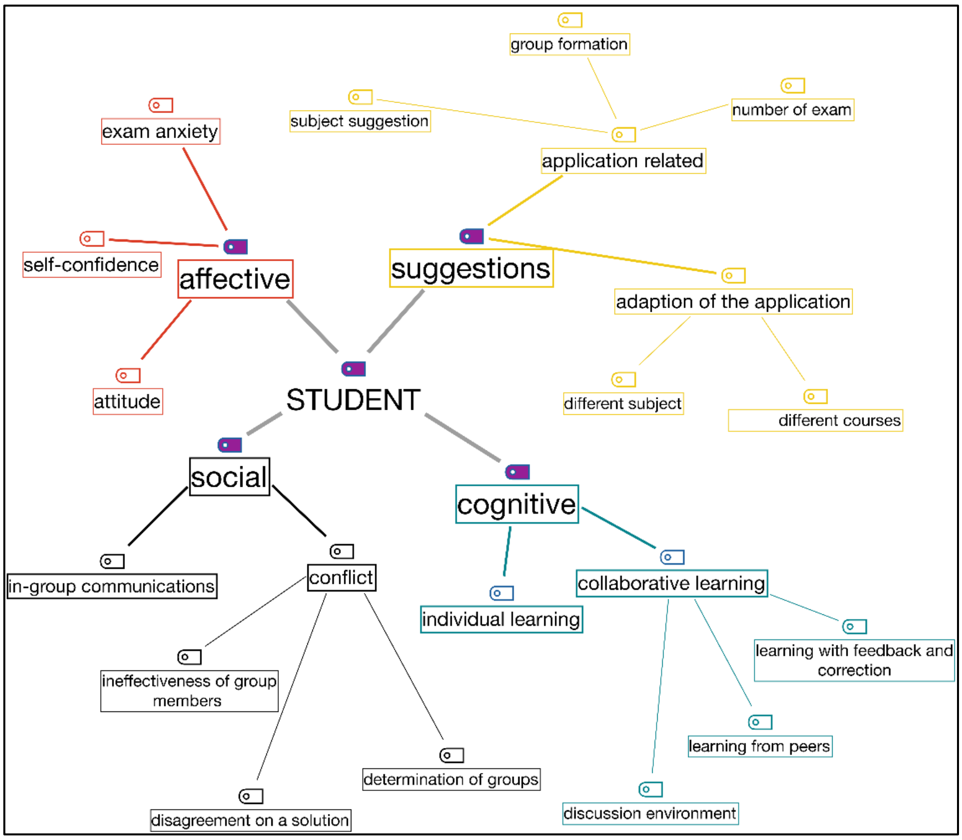


Figure 3. The themes and sub-themes determined based on student views

The students stated that they had the opportunity to discuss with their peers and collaborated with them to accomplish individual learning. The students claimed that they experienced self-learning in algebraic expressions, and the technique improved their classroom performances. The dialogue between the student S2 and the researcher is presented below.

> ***R:*** *[...]OK, for example, did you feel that your performance in algebraic expressions improved in the lesson? Were there times where you could say "I went up to the board more, I raised my hand more"? (08:10-08:21)*
> ***S2:*** *I was not doing anything in the first semester. I used to sit in my desk and not being active. (08:21-08:26)*
> ***R:*** *The second semester? (08:27-08:27)*
> ***S2:*** *I tried to improve a little in the second semester; then you came, and that was nice, I learned better. I was learning from our teacher, my friends and you. (08:28-08:38)*
> ***R:*** *Did this activity contributed to your improvement? Let's say, for increasing your classroom performance? (08:39-08:46)*
> ***S2:*** *Of course, it did. (08:46-08:46)*

Similarly, other students emphasized the contribution of the collaborative testing as follows: "*I understood similar terms better*", "*You arrived, and I understood the terms coefficient and constant*", "*I did not know the term numbers very well, I did not understand, I understood them during the collaborative tests. Now I can do them much better*". Furthermore, certain students stated the impact of the collaborative testing on learning very clearly. For example, participant S1 stated: "*I learned more with collaborative tests*". Also, during the application, it was observed that the students solved the algebra problems easily through discussions with their peers. These student opinions were categorized under the theme "collaborative learning". As seen in Figure 4, student views on collaborative learning were categorized in three groups: "learning from peers", "discussion environment" and "learning with feedback and correction. Students stated that they understood the question better and got better grades when they solved the test with their peers. Participant S13 stated that they solved the problems in solidarity and collaboratively: "*It was easier when we did it together; because when we combined our ideas, great things emerged*." The conversation between the researcher and S12 is presented below:

> ***S12:** I had not really understood the algebraic expressions that much, but when we worked in a group, we learned by sharing our ideas. (4.06-4.15)*
> ***R:** So, why do you think that you understood better in a group? (4.15-4.19)*
> ***S12:** When compared to individual work, for example, my friends know the topics that I do not know and they help, in individual study, I cannot continue when I cannot remember but I comprehend better when I am in a group. (4.19-4.33)*

The students stated that it was easier to work with their peers and solve the problems they could not solve by themselves. For example, the participant S2 stated the following: "*For example, my friends can solve something that I cannot, I can solve what they cannot. If my friends are smarter than me, I can learn things I did not know from them. If a friend does not know what I know, then I can teach them.*" Thus, the students stated that they learned from their peers in the collaborative testing. The students mentioned that they experienced difficulties in writing verbal algebra expressions, but they learned it with the help of their groupmates. Also, some students claimed that they better understood multiplication of a binomial algebraic expression by a fixed number better after their peers explained it. Certain students experienced difficulties with coefficients in algebraic expressions, but they overcame this difficulty with the help of their peers: "[…] *coefficients, for example, I could not understand, but S1 informed us how to solve it*". On the other hand, the fact that students solved the algebra problems with group discussions demonstrated the contribution of the exercise to cognitive learning. S3 preferred the collaborative testing for the following reasons: "[…] *because we can discuss them, but we cannot discuss anything in individual tests, we make mistakes because we do not understand some questions*". S2 aslo stated that the algebra problems were solved by discussion: "*Our friends who understood would explain it to us, and we solve it by discussing*."

The "feedback" stage was significant in student statements on collaborative testing exercises. The feedback provided by the researchers on the misunderstandings, mistakes and shortcomings in student responses after each test to ensure the elimination of the misunderstandings and to correct mistakes or to ensure a shortcoming was remedied was also reflected in student opinions. Based on these opinions, the final sub-theme "learning with feedback-correction" was determined in the "learning with collaboration" theme. S12 stated that collaboration with the researchers was beneficial and mentioned their acquisitions due to this collaboration as follows: "*For example, we made mistakes in certain problems; we corrected these, which improved our grades.*" On the other hand, S10 stated how they overcame the difficulties within the group: "*We made mistakes in the first test, we corrected these in the second test, and since we understood the problem, we did not make the same mistake again*." It was observed that the feedback provided by the researchers on algebra test questions was employed by the students in the next test, leading to higher grades. For example, S8 stated that they had the opportunity to correct their mistakes thanks to the feedback:

> ***R:** We talked about your mistakes and problems you experienced after each test with the group. Did this have an affect your performances in the next test? (01:44-01:58)*
> ***S8:** Yes. (01:58-01:59)*
> ***R:** Did it? Can you elaborate on these affects? (01:59-02:23)*
> ***S8:** For example, I experienced problems in tables. We experienced problems in tables in the first test and our grades were bad. In the second, we understood the tables and could identify the general rule. (02:25-03:04)*
> ***R:** I understand, anything else? Did you mean the tables in algebraic expressions? (03:05-03:12)*
> ***S8:** Yes. These, and there was 5n+1, we learned the required steps. We did not know these. (03:12-03:21)*

Similarly, it was observed that the students learned the parts where they experienced difficulties in algebraic expressions through feedback. Also, some students emphasized the significance of feedback in modeling algebraic expressions.

It was observed that the student opinions on learning algebraic expressions with collaborative testing were not only associated with "cognitive" learning but also "affective" learning. These views were categorized under three themes: "exam anxiety", "attitude" and "self-confidence". Certain students stated that their exam anxiety decreased after collaborative testing activities, while others stated that they were still very anxious after collaborative testing. S11 stated that the exercise reduced their exam anxiety: "*I was quite anxious about getting bad grades before, but I am not like that anymore, I do not feel very anxious*." Similarly, the conversation between the researcher, S6 revealed that the exercise reduced exam anxiety:

> ***R:*** *You are anxious in individual tests, but did your anxiety levels change after the collaborative testing? (4.33-4.40)*
> ***S6:*** *In fact, I am not anxious in collaborative tests, but I am very anxious in individual tests. (4.40-4.49)*
> ***R:*** *Ok, did you experience anxiety in the individual tests after the application? (4.49-4.53)*
> ***S6:*** *Yes, but less than before. (4.53-4.56)*

It was observed that the activities also affected student attitudes towards mathematics. Most students stated that their attitudes towards mathematics improved and that enjoyed the course more. The dialogue between S9 and the researcher reflects this:

> […]
> ***S9:*** *I used to think that mathematics was boring, but now I think its fun. (03:54-03:58)*
> ***R:*** *Why do you consider it fun now? (04:11-04:13)*
> ***S9:*** *Mathematics is involved in everything. (04:13-04:18)*
> ***R:*** *OK, what changed after the collaborative tests and made you think maths is fun? (04:19-04:23)*
> ***S9:*** *We learn everything in math. For example, algebraic expressions, addition; for example we add and measure areas everywhere. (04:26-04:37)*
> ***R:*** *Yes, but why is it the case after collaborative testing? Can you elaborate with examples? (04:37-04:56)*
> ***S9:*** *An example? Well, when you first arrived, in the first collaborative test, I could not do anything, I looked at the problems, and then, it started to be fun, and now I fully participate in finding solutions. (04:56-05:14)*

Similarly, it was found that the attitudes of the other students towards mathematics changed after the collaborative testing due to the algebraic attainments with which they experienced difficulties before. S8, one of the two students who stated that their self-confidence increased after the exercise, stated the following: "*It got better. How can I explain it? My confidence in mathematics improved*." S3 stated the following: "*We had less confidence in individual tests, but we had a lot of confidence in collaborative tests*". However, it was observed that one student (S13) still considered mathematics difficult, even after the exercise. On the other hand, S3 liked mathematics before the exercise and stated that there was no change in this positive feeling about the subject. Only one student (S1) stated that (s)he did not like mathematics, and this did not change, as this was a difficult topic area.

Students also expressed opinions on communication within the groups in the sessions were their opinions were sought after the collaborative testing exercise. These views were categorized in the "social" theme as seen in Figure 1. The students stated that the exercise had a positive effect on the relations with their peers. S11 stated the following: "*We did not talk to her/him much, there was a distance between us, we never played games together. Now, we always play together. We include her/him in our games*." S12 stated the following: "*We were not close, when compared to other friends, but now I am on good terms with her/him*". S7 stated that (s)he became very friendly with a groupmate and they play games outside of school: "*Our relationship was not good before, then, it became better. We started to play together after school. So, it improved my relationship with her/him*." On the other hand, some students stated that social communication with their peers remained the same and the exercise did not lead to any changes: "*It was already good. I know S2 from folk dancing sessions; yes, we were together in folk dancing. My relationship was not good with S1, and it is still the same, no change.*" S1 similarly confirmed that communication with the groupmates remained the same.

However, certain students expressed negative opinions. These views were categorized in the "conflict" theme that included "disagreement on a solution", "ineffectiveness of group members" and "determination of groups" sub-themes. Some students mentioned the aspects of collaborative testing that they disliked and stated that sometimes the discussions did not lead anywhere, certain students who disagreed with a solution insisted on their own solution, which made it difficult for their group to complete the test. S7 stated the following: "*When solving the problems, someone solved the problem and everyone was offering their own solution, but that person insisted on their own solution, leading to problems.*" Other students stated that their groupmates did not share responsibilities. The opinions of the students who complained about non-participation of certain group members, who were interested in doing other things and did contribute to the solution were categorized under the ""ineffectiveness of group members" theme. S1 stated the following: "*It also had disadvantages. For example, lets say that someone is very smart and someone else is not. Only the smart one works*." Another student, S4 assessed this disadvantage based on time and gender, a significant finding of the study. The comment of S4 is as follows: "*When there is no consensus, they just say 'I will not do it' and they immediately withdraw, and we lose time while others in the group try to convince this person, and when the group includes both boys and girls, the boys exclude the girls, and tell us you do it, and so it remains on us to do it."* Also, most students who expressed negative opinions argued that they did not like their groups and wanted to change their groups. A few students stated that they were not happy with their groupmates, and they would be more successful if they were in a different group.

The students also expressed recommendations to improve the exercise. The analysis of the students views on recommendations about the application revealed "subject suggestion", "group formation" and "number of exam" themes. Certain students stated that the topic of algebraic expressions was quite difficult and suggested that the application should be conducted on another topic. For example, the participant coded S10 stated that "*It could be in mathematics, but I think the topic should be changed. Algebra is very difficult*." Similarly, S14 stated that "*Fractions are easier but algebraic expressions are difficult..*." Furthermore, some students expressed negative attitudes towards group formation and argued that they wanted to be in a group with intelligent students. They suggested that the groups should be formed from intelligent students. S1 stated that it was unfair: "*I wish my group had a smart one too, then I would get 100 in oral test. I mean, I cannot say it for all the groups, but 2-3 groups had the smart students, so that was why…*" Also, some students stated that they should have determined the groups. On the other hand, it was determined that the individual tests were more effective, compared to collaborative tests according to the student suggestions. For example, S5 stated the following: "*We can use different things for this. For example, what can happen in individual tests? More help could be available, the number of these could be less, the number of collaborative tests could be more. Because they can do the test collaboratively, but they could not do it in the individual test*."

Most students stated that this practice could be conducted both in mathematics and other subjects. They stated that they comprehended the problems better and got better grades when they solved the problems together; and thus, they wanted the practice to be adapted to other subject lessons. These student recommendations were categorized in the "repetition of the application" sub-theme.

The student opinions that the application could be adopted to other courses were significant. Thus, these opinions were included under the theme of "different subjects", and it was seen that the students wanted the application to be adapted in Turkish, English, Science, and Physical Education subjects. The students who wanted the adoption of the application in different subjects stated that it should be based on the difficulty of the lesson. It was determined that one of the reasons for suggesting to adapt the practice in the science subject was the difficulty of the science subject based on student opinions. S7 stated that the application should be adopted in the English subject as well: "*Because we learn a different language in the subject, if it is adopted, we will learn English easily. I think it would be nice if it was implemented in English.*"

4. DISCUSSION and CONCLUSION

As mathematics education experts are becoming increasingly convinced that students could learn with communication, and peer-to-peer communication is considered as a primary learning tool to be included in the curriculum of several countries (Campbell, 2021). It is known that collaborative learning activities designed with an effective and organized approach provide significant opportunities for the construction of student knowledge. On the other hand, the idea that assessment could also be conducted with groupwork and that could be employed as a means for permanent learning in tests have been discussed (Çalık-Uzun & Birişçi, 2018). This idea was the foundation of this study, which aimed to employ collaborative testing as an alternative measurement and assessment method and to investigate its impact on mathematical learning outcomes. At the same time in the present study, students' views were also obtained on collaborative testing, and their experiences in the process were ascertained.

The analysis of the findings on the impact of collaborative testing, employed as an alternative assessment instrument, on cognitive learning outcomes revealed that the mean experimental group score was higher than that of the students in the control group; however, the difference was not statistically significant. Similarly, LoGiudice, Heisz, and Kim (2021) reported that student perceptions about collaborative testing were generally positive; however, they did not find evidence of a difference between the post-test grades that favored collaborative testing. Similarly, the qualitative findings of the study demonstrated that the student opinions were positive; they gained a better understanding when solving the problems in collaboration, which also improved their performances. The collaborative testing allowed students to discuss the concepts with their peers and understand them instead of experiencing disappointment with test results or not completing test items when in doubt (LoGiudice, Heisz & Kim, 2021). Various studies reported that collaborative testing increased the interaction between students (Giraud, 1997; Ioannou & Artino, 2010; Keller & Steinhorst, 1995; Magel, 1998) and contributed to learning (Bloom, 2009; Breedlove et al., 2007; Guiliodori et al., 2008; Muir & Tracy, 1999; Zimbardo et al., 2003; Zipp, 2007).

It was determined that collaborative testing had a positive impact on affective learning of students. Spielberger, Anton, & Bedell (1976) defined exam anxiety that is as an individual's disposition to worry and have interfering thoughts, feel mental confusion, and tension and give a physical reaction during any exam (cited in Alibak, Talebi, Neshat-Doost, 2019, p.2). Exam anxiety has two dimensions as worry and emotionality. In this study, the exam anxiety levels of the students were investigated as an affective learning variable, and it was determined that the collaborative testing decreased total exam anxiety score. However, the impact was not statistically significant. In a study conducted by Breedlove et al (2004), no significant difference was determined between exam anxiety in collaborative and individual tests, and it was reported that the exam anxiety was effective in organization of knowledge and the effect could vary based on the testing approach. On the other hand, when test anxiety score results are evaluated according to sub-dimensions, it was found that worry sub-dimension scores increased whereas emotionality sub-dimension scores decreased. The worry dimension is the cognitive aspect of test anxiety and includes the individual's negative evaluations of himself in general, negative thoughts about his failure, and incompetence. The emotionality dimension is the stimulation of the autonomic nervous system, which constitutes the sensory physiological aspect of test anxiety. Physical experiences such as rapid heartbeat, chills, perspiration, nausea, redness-yellowing, irritability and tension are symptoms of emotionality (Öner, 1990: 1). Therefore, it can be said that collaborative testing has no effect on the negative cognitive thoughts that students feel during any exam. This result is also supported by the qualitative findings of the study.

Numerous studies have noted that the cooperation of the students in completing the collaborative testing and the success they attained as a result of this cooperation play a significant role in boosting the students' self-confidence (Breedlove et al., 2004; Dalmer, 2004; Grubb, 2014; Mahoney, 2019). In the current study, it was also determined in the opinions of some of the participants that the improved self-confidence in the collaborative testing process contributed to the reduction in test anxiety experienced, and the students stated that their anxiety about individual exams decreased. This outcome backs up a number of research in the literature (Amaral, 2004; Mahoney & Reeves, 2019; Pandey & Kapitanoff, 2011; Willard, 2015; Zimbardo et al., 2003). Students' ideas expressing the contrary of these thoughts are also found when the qualitative study findings are reviewed. In fact, some students claimed that when working on solutions with others in their group, they had no worries, but taking the exam alone, they felt quite uncomfortable. It can be said that these students depend on the ideas of their other group members to solve the problems together, and they are nervous since they believe that other students won't be able to help them with individual tests.

The mean attitude towards mathematics scores of the students, an affective variable, increased in the post-test in comparison to the pre-test. Thus, it can be stated that collaborative testing was effective on mathematical attitudes, since education plays a key role in changes in attitude (Duatepe & Çilesiz, 1999). Furthermore, the qualitative study findings demonstrated that most students had fun when discussing the test activities within a group and enjoyed talking about mathematics. However, it was observed that the attitudes of some students who already liked mathematics or those who never liked it, did not change. This could be explained by student beliefs associated with learning mathematics.

The qualitative study findings on the opinions of students on collaborative testing suggested that collaborative testing improved student attitudes towards mathematics as a subject, their cognitive and social communications in the group, and classroom performance (Çalık-Uzun & Birişçi, 2018). It was revealed that they would be happy to continue the practice in future lessons. Some students expressed negative opinions, claiming that they were ignored by their groupmates; they were dissatisfied with their peers who did not contribute to collaboration and mentioned that there was occasionally discussion which were inconclusive. Some studies reported that certain limitations of group members could lead to obstacles to collaborative testing. The disadvantage of the technique is the fact that the inactive group members during the collaborative tests would also benefit from the success of the group in the test environment (Çalık-Uzun & Birişçi, 2018).

Our study supports the results of many studies in the literature by showing that collaborative testing can be usefully applied in classroom exams. It was determined that collaborative testing contributed to the cognitive and affective learning of students, as supported by the students' opinions. It was determined that there was no statistically significant difference between several variables that were analyzed in the study. However, it was observed that total variable scores increased, which was also supported by the qualitative findings. It could be suggested that longitudinal studies could be conducted for collaborative testing. Because students may require a certain period of time to adapt to the transformation from a traditional measurement and assessment system to a collaborative testing method. The current study presented a cross-section of the assessment of algebra learning area attainments. Since the students were in the transition period from a pre-algebraic point to the algebraic period, it was not surprising that there were no significant differences between the cognitive learning scores. While investigating the effect of the study on cognitive learning, the algebra attainment test could not be administered as a pre-test because the students encountered algebraic expressions for the first time in 6th grade. This situation caused us to limit the study to a simple experimental design. Different results can be obtained with quasi-experimental studies designed for other learning domains.

It could be suggested that problems in the solution process could be eliminated with the inclusion of collaborative testing activities in assessment as an alternative to traditional individual assessment techniques (Johnson & Johnson, 1999). Thus, the repetition of the study on different topics and levels would contribute to the literature.

In summary, we repeat the call by Muir and Tracy (1999) made over 2 decades ago and invite teachers to try collaborative testing to assess their students' cognitive and affective learning in mathematics lessons. As the researchers of this study, we defend that it is valuable to use collaborative testing in math courses as an alternative assessment tool.

Acknowledgement

The researchers confirmed that the data in this study were collected before the year 2020.

The research was funded by the Artvin Çoruh University, Scientific Research Projects Coordination Unit with Project number 2016.S34.02.04. We would like to thank Artvin Çoruh University, Scientific Research Projects Coordination Unit for their support.

The authors wish to thank Assoc. Prof. Dr Salih Birişçi for his comments on the design and his support at every stage of the study and the students and mathematics teacher who participated in the current study.

5. REFERENCES

Alibak, M., Talebi, H., & Neshat-Doost, H. T. (2019). Development and validation of a test anxiety inventory for online learning students. *Journal of Educators Online*, *16*(2).

Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning*, *74*, 5–15 https://doi.org/10.1002/tl.7401

Amaral, K. E. (2004). *Collaborative learning and testing in introductory general chemistry.* [Unpublished doctoral dissertation]. Florida University, ABD.

Aşkar, P. (1986). Matematik dersine yönelik tutumu ölçen likert tipi bir ölçeğin geliştirilmesi [Developing a likert-type scale that measures attitude towards the mathematics lesson]. *Education and Science*, *11*(62), 31-36.

Baki, A., & Birgin, O. (2002, September). *Matematik eğitiminde alternatif bir değerlendirme olarak bireysel gelişim dosyasi uygulamasi* [*Individual development file application as an alternative assessment in mathematics education*]. V.Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, 16-18 Eylül, Ankara.

Baki, A., & Birgin, O. (2004). Reflections of using computer-based portfolios as an alternative assessment tools: A case study. *Turkish Online Journal of Educational Technology*, *3*(3), 11.

Başkale, H. (2016). Nitel Araştırmalarda Geçerlik, Güvenirlik ve Örneklem Büyüklüğünün Belirlenmesi [Determination of validity, reliability and sample size in qualitative studies].*Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi*, *9* (1), 23-28. Retrieved from https://dergipark.org.tr/en/pub/deuhfed/issue/46796/586804

Bloom, D. (2009). Collaborative test taking: Benefits for learning and retention. *College Teaching* *57*(4), 216–20. https://doi.org/10.1080/87567550903218646

Breedlove, W., Burkett, T., & Winfield, I. (2004). Collaborative testing and test performance. *Academic Exchange Quarterly*, *8*(3), 36-40.

Breedlove, W., Burkett, T., & Winfield, I. (2007). Collaborative testing, gender, learning styles, and test performance. *Int J Scholarsh Teach Learn*, *4*(2), 33-42.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., Demirel, F. (2020). *Eğitimde bilimsel araştırma yöntemleri. (29. Edition).* Ankara: Pegem Akademi Publishing.

Campbell, T. G. (2021). An examination of discourse analytic methods in the context of mathematical group work. *International Journal of Mathematical Education in Science and Technology*, 1-24. https://doi.org/10.1080/0020739X.2021.1944681

Cansız-Aktaş, M. (2018). Öğretmenlerin performans görevlerine bakışları: hizmet içi eğitim almış zümre başkanları örneği [Teachers' views on the performance tasks: a case of department heads who had in-service training] *Journal of Computer and Education Research, 6*(12), 320-341. https://doi.org/10.18009/jcer.465378

Cantwell, E.R., Sousou, J., Jadotte, Y.T., Pierce, J., & Akioyamen, L.E. (2017). Collaborative testing for improving student learning outcomes and test taking performance in higher education: A systematic review. *Campbell Systematic Reviews, 13* (1), 1-18. https://doi.org/10.1002/CL2.186

Creswell, W.J. (2016). *Araştırma deseni: Nitel, nicel ve karma yöntem yaklaşımları* [Research design: Qualitative, quantitative and mixed methods approaches]. (S. Beşir Demir, Trans.). Ankara: Eğiten Kitap Press (Original work published 2013).

Çalık-Uzun, S., & Birişçi, S. (2018). Reflections on collaborative testing: student and teacher opinions. *Trakya University Journal of the Faculty of Education, 8*(3), 591-612. https://doi:10.24315/ trkefd.352717.

Çalışkan, H., & Kaşıkçı, Y. (2010). The application of traditional and alternative assessment and evaluation tools by teachers in social studies. *Procedia-Social and Behavioral Sciences, 2*(2), 4152-4156. https://doi.org/10.1016/j.sbspro.2010.03.656

Çepni, S. (2009). *Araştırma ve proje çalışmalarına giriş* [Introduction to research and project studies]. *(4. Edition)* Trabzon: Celepler Press.

Çoruhlu, T.Ş., Nas, S. E., & Çepni, S. (2009). Fen ve teknoloji öğretmenlerinin alternatif ölçme değerlendirme tekniklerini kullanmada karşılaştıkları problemler: Trabzon Örneği [Problems facing science and technology teachers using alternative assessment tecnics: Trabzon sample]. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, *(1)*1, 122-141. https://dergipark.org.tr/tr/pub/yyuefd/issue/13711/165998

Dalmer, D. (2016). Collaborative test taking with adult learners. *Adult Learning, 15*(3-4), 4-7

Duatepe, A., & Çilesiz, Ş. (1999). Matematik tutum ölçeği geliştirilmesi [Developing a math attitude scale]. *Hacettepe University Faculty of Education Journal 16(*16). https://dergipark.org.tr/en/download/article-file/88087

Duban, N., & Küçükyılmaz, E.A. (2008). Primary education pre-service teachers' opinions regarding to the use of alternative measurement-evaluation methods and techniques in practice schools. *Elementary Education Online, 7*(3), 769-784. https://dergipark.org.tr/en/download/article-file/90936

Dursun, Ş., & Bindak, R. (2011). İlköğretim II. kademe öğrencilerinin matematik kaygılarının incelenmesi [Primary education II. examining the mathematics anxiety of secondary school students]. *Cumhuriyet University Journal of Social Sciences, 35*(1), 18-21. http://cujos.cumhuriyet.edu.tr/en/download/article-file/49853

Gelbal, S. & Kelecioğlu H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar [Efficacy perceptions of teachers about measurement and evaluation methods and the problems they face]. *Journal of Hacettepe University Faculty of Education*, 33, 135-145. https://dergipark.org.tr/tr/download/article-file/87640

Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education, 5*(3), 1. https://doi.org/10.1080/10691898.1997.11910598

Giraud, G., & Enders, C. (2000, April). The effects of repeated cooperative testing in an introductory statistics course. *Annual Meeting of the American Educational Research Association*. April, San Diego, CA.

Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2008). Collaborative group testing benefits high-and low-performing students. *Advances in Physiology Education, 32*(4), 274-278. https://doi.org/10.1152/advan.00101.2007.

Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the behavioral sciences* (9th ed). London: Thomson Wadsworth Publishing.

Grubb, A. G. (2014). The effects of collaborative testing and the testing effect on student achievement and confidence of undergraduate business students. [Unpublished doctoral dissertation]. South Alabama University, ABD.

Hendrickson, J.M., Brady, M.P. & Algozzine, B. (1987). Peer-Mediated testing: The effects of an alternative testing procedure in higher education. *Educational and Psychological Research 7*(2), 91-101.

Hodges, L. C. (2004). Group exams in science courses. *New Directions for Teaching and Learning, 100*, 89-93. https://doi.org/10.1002/tl.175

Ioannou, A., & Artino, A. R, Jr. (2010). Learn more, stress less: exploring the benefits of collaborative assessment. *PHD College Student Journal, 44*(1), 189–199.

Johnson, D. W., & Johnson, R. T. (1999). Cooperative learning and assessment. In Kluge, D., McGuire, S., Johnson, D., Johnson, R. (Eds.), *JALT applied materials*: *Cooperative Learning*. (pp. 163-178). Japan Association for Language Teaching.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Kapitanoff, S., & Pandey, C. (2018). Collaborative testing in statistics: Group interaction, anxiety, and class performance. *Statistics Education Research Journal*, *17*(2), 51-67. https://doi.org/10.52041/serj.v17i2.158

Karamustafaoğlu, S., Çağlak, A., & Meşeci, B. (2012). Alternatif ölçme değerlendirme araçlarına ilişkin sınıf öğretmenlerinin öz yeterlilikleri. [Self-efficiency of primary school teachers related to the alternative testing and assessment tools] Amasya Üniversitesi Eğitim Fakültesi Dergisi, 1(2), 167-179. https://dergipark.org.tr/tr/pub/amauefd/issue/1727/21168

Keller, C. M., & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, *3*(2). https://doi.org/10.1080/10691898.1995.11910485

Kuran, K., & Kanatlı, F. (2010). The evaluation of classroom teachers' opinions on the alternative assessments techniques. *Journal of Mustafa Kemal University Institute of Social Sciences*, 6(12), 209-234. https://dergipark.org.tr/tr/pub/mkusbed/issue/19557/208441

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20(3), 975-978.

LoGiudice, A. B., Heisz, J. J., & Kim, J. A. (2021). Does collaborative testing in the classroom enhance delayed transfer of knowledge? *Scholarship of Teaching and Learning in Psychology.* Advance online publication. https://doi.org/10.1037/stl0000267

Lusk, M., & Conklin, L. (2003). Collaborative learning to promote learning. *Journal of Nursing Education*, *42*(3), 121-124. https://doi.org/10.3928/0148-4834-20030301-07

Magel, R. C. (1998). Using cooperative learning in a large ıntroductory statistics class. *Journal of Statistics Education* [Online], 6(3). https://doi.org/10.1080/10691898.1998.11910621

Mahoney, J. W., & Harris-Reeves, B. (2019). The effects of collaborative testing on higher order thinking: Do the bright get brighter? *Active Learning in Higher Education*, *20*(1), 25–37. https://doi.org/10.1177/1469787417723243

Mitchell, N., & Melton, S. (2003). Collaborative testing: An innovative approach to test taking. *Nurse Educator*, *28*(2), 95-97. https://doi.org/10.1097/00006223-200303000-00013

Muir, S. P., & Tracy, D. M. (1999). Collaborative essay testing: just try it! *College Teaching, 47*(1), 33-35. https://doi.org/10.1080/87567559909596077

Öner, N. (1990). *Sınav kaygısı envanteri el kitabı* [Test anxiety inventory handbook]. İstanbul: Yöret Press.

Pandey, C., & Kapitanoff, S. (2011). The influence of anxiety and quality of interaction on collaborative test performance. *Active Learning in Higher Education*, 12(3), 163-174

Pituch, K. A., & Stevens, J. P. (2009). 434 Binary Logistic Regression. In Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS (pp. 434-470). Routledge.

Plano Clark, V.L., & Creswell, J. W. (2008). *Mixed methods reader*. Sage Publications.

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in physiology education*, *26*(1), 37-41.

Reyes, M. D., & Castillo, A. C. (2015). Test anxiety and college students' performance on mathematics departmental examination: basis for mathematics achievement enhancement. *Asia Pacific Journal of Education, Arts and Sciences*, *2*(1), 62-69.

Quarstein, V. A., & Peterson, P. A. (2001). Assessment of cooperative learning: A goal-criterion approach. *Innovative Higher Education, 16*, 59-77.

Schutz, P. A., Distefano, C., Benson, J., & Davis, H. A. (2004). The emotional regulation during test-taking scale. *Anxiety, Stress & Coping, 17*(3), 253-269.

Slusser R,, & Erickson R.J. (2006). Group quizzes: an extension of the collaborative learning process. *Teach Sociol, 34*(3), 249–262.

Spielberger, C. D. (Ed.). (1995). *Stress and emotion* (Vol. 15). Taylor & Francis.

Toptaş, V. (2011). Classroom teachers' perceptions about the use of alternative assessment and evaluation methods in mathematics courses. *Education and Science, 36*(159), 205-219.

Willard, C. (2015). Effects of collaborative reasoning on students mathematics performance and numerical reasoning abilities. [Unpublished doctoral dissertation]. Temple University, ABD.

Yenilmez, K., Girginer, N., & Uzun, Ö. (2007). Mathematics anxiety and attitude level of students of the faculty of economics and business administrator: The Türkiye model. *International Mathematics Forum, 2*(41), 1997-2021.

Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri* (6th ed.) [Qualitative research methods in the social sciences]. Seçkin Press.

Yuretich, R.F., Khan, S.A., Leckie, R.M., & Clement, J.J. (2001). Active-learning methods to improve student performance and scientific interest in a large introductory oceanography Course. *Journal of Geoscience Education, 49*(2), 111-119.

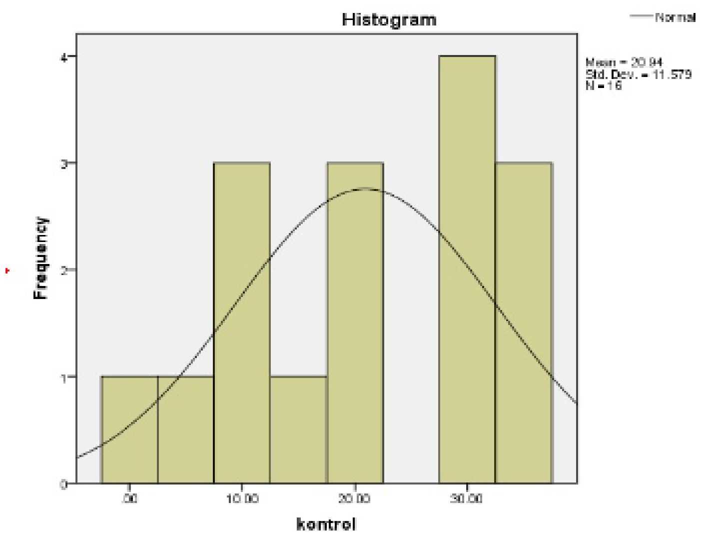
Zimbardo, P. G., Butler, L. D., & Wolfe, V. A. (2003). Cooperative college examinations: More gain, less pain when students share information and grades. *The Journal of Experimental Education, 71*, 101-125.

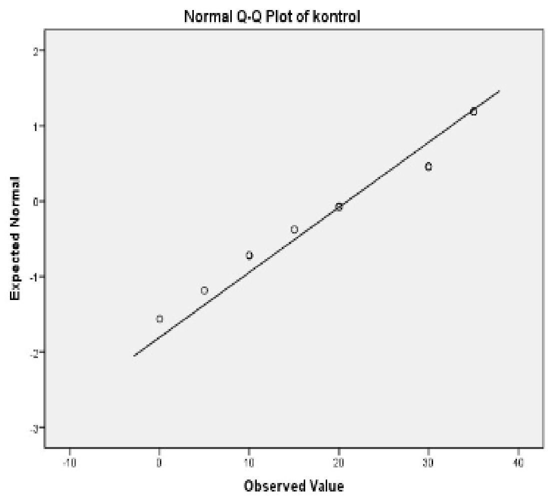
Zipp, J. F. (2007). Learning by exams: The impact of two-stage cooperative tests. *Teaching Sociology, 35*, 62-76.

Appendix -A

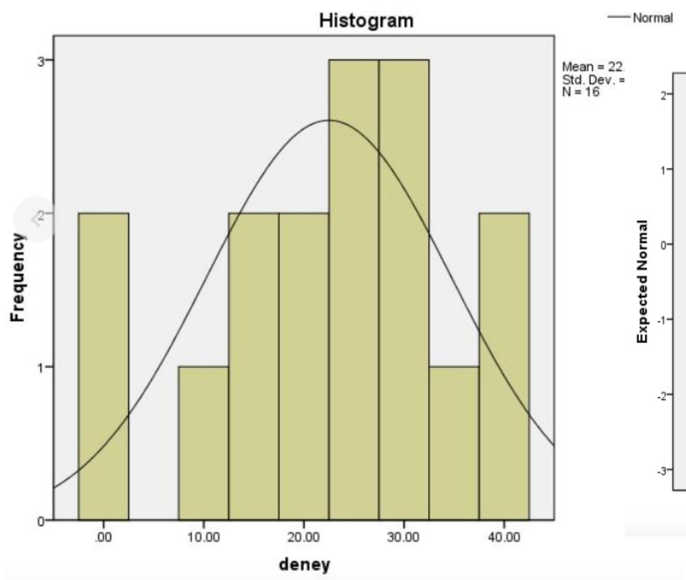
Histogram and Q-Q plot graph of the control group

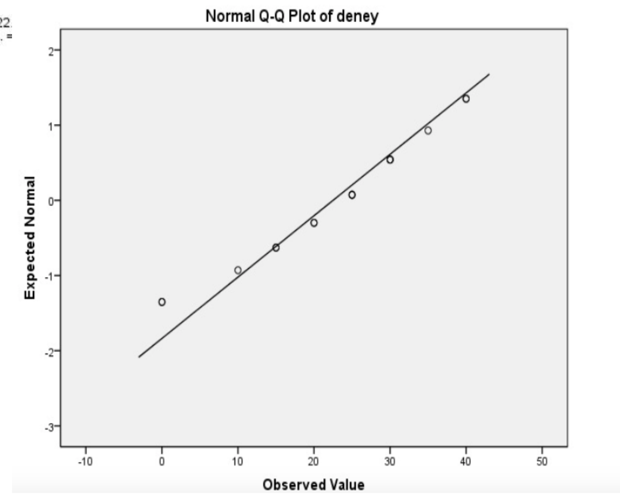
kontrol





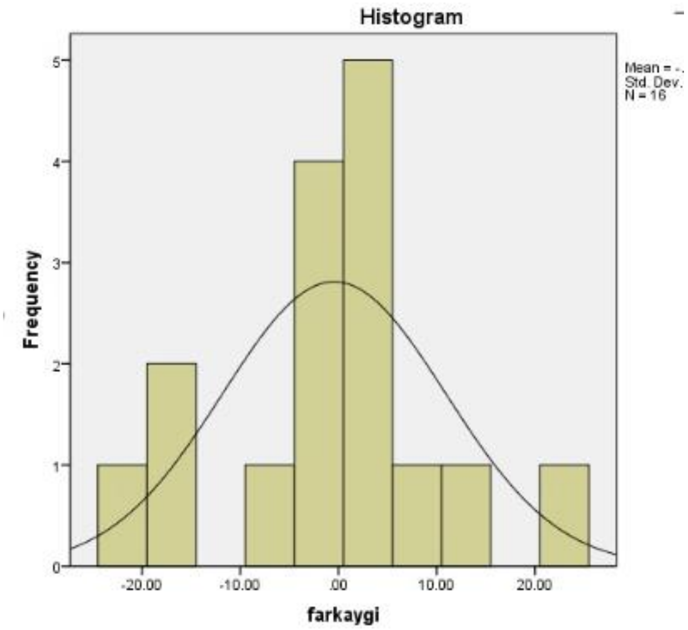
Histogram and Q-Q plot graph of the experimental group

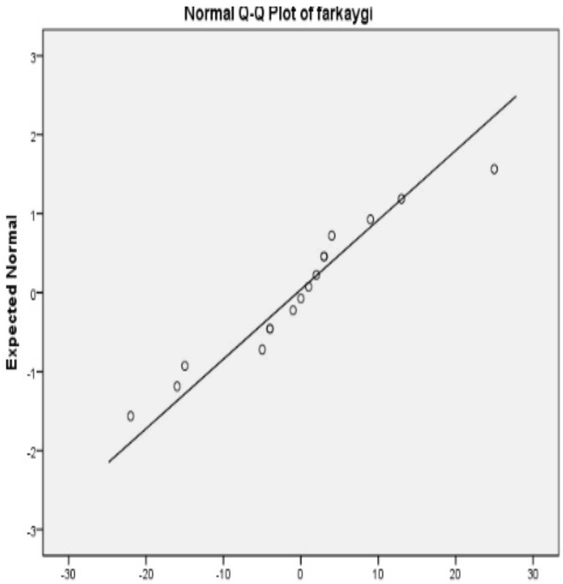




Appendix -B

Histogram and Q-Q plot of the normal distribution of exam anxiety data





Appendix -C

Histogram and Q-Q plot of the normal distribution of mathematical attitude da