IJPES
www.ijpes.com
International Journal of Psychology and Educational Studies
ISSN: 2148-9378

Measuring Test Anxiety Using Physiological Markers: A Comparative Study of Multiple-Choice and Open-Ended Tests*

Meltem ACAR GÜVENDİR[1], Seda DONAT BACIOĞLU[2], Hasan ÖZGÜR[3], Sefa UYANIK[4], Fatmagül GÜRBÜZ AKÇAY[5], Emre GÜVENDİR[6]

1Faculty of Education, Trakya University, Edirne, Türkiye 0000-0002-3847-0724
2 Faculty of Education, Trakya University, Edirne, Türkiye 0000-0001-9901-0601
3 Faculty of Education, Trakya University, Edirne, Türkiye 0000-0002-8035-0320
4 Fatih Faculty of Education, Trabzon University, Türkiye 0000-0002-7468-0067
5 Faculty of Education, Trakya University, Edirne, Türkiye 0000-0001-9101-5889
6 Faculty of Education, Trakya University, Edirne, Türkiye 0000-0003-1226-9878

ARTICLE INFO

Article History
Received 30.08.2024
Received in revised form 02.01.2025
Accepted 02.02.2025
Article Type: Research Article

CC BY NC ND

ABSTRACT

Different types of test items influence students' test anxiety, and physiological measures such as heart rate provide a means of measuring this anxiety. This study aimed to explore the connection between test anxiety and examination item formats. It centered on 20 junior university students in Western Türkiye. The research monitored students' heart rate changes during assessments, encompassing both multiple-choice and open-ended formats. Using a relational survey model, the study focused on statistically significant differences in mean rank scores for students' state and trait-based test anxiety, assessed through the Friedman test. Additionally, it examined heart rate variance, a physiological stress indicator, shedding light on heart rate fluctuations experienced by students during and outside exams. Participants exhibited higher levels of trait test anxiety. Moreover, heart rate changes during multiple-choice items exceeded those during open-ended items. These findings imply that multiple-choice items may trigger elevated anxiety in students, whereas open-ended items may elicit less anxiety, encouraging flexible thinking skills utilization.

Keywords:
Test anxiety, item type, heart rate, variance, measurement, evaluation

1. Introduction

Anxiety is defined as a state of worry and unease about a subjective situation that may or may not happen in the present or future, or may even be unlikely to occur (Lazarus & Folkman, 1984; Şahin, 2019). Spielberger et al. (1968) posit that the crux of anxiety lies in an individual's subjective interpretation of environmental stimuli as personal threats. Spielberger (1966) proposed that anxiety comprises two distinct factors: state anxiety and trait anxiety. A substantial distinction has been identified between state anxiety and trait anxiety (Cattell & Scheier, 1961; Spielberger, 1980). State anxiety is defined as a transient emotional state, which is more associated with the activation or mobilisation of the autonomic nervous system. It is characterised by the conscious experience of fear and tension/excitement. Conversely, trait anxiety is considered a component of an individual's personality. Individuals with chronic anxiety tend to perceive a wide range of conditions, physically or psychologically, as threatening even when they are not genuinely dangerous (Kapur et al., 2019).

*This study was produced as part of project 2021/05, funded by the Trakya University Research Project Fund (TUBAP).
Corresponding author's address: Faculty of Education, Trakya University, Edirne, Türkiye
e-mail: meltemacar@trakya.edu.tr
Citation: Acar Güvendir, M., Donat Bacıoğlu, S., Özgür, H., Uyanık, S., Gürbüz Akçay, F., & Güvendir, E. (2025). Measuring test anxiety using physiological markers: A comparative study of multiple-choice and open-ended tests. *International Journal of Psychology and Educational Studies, 12*(2), 154-168. https://dx.doi.org/10.52380/ijpes.2025.12.2.1392

As is evident in the extant literature, anxiety is a recurrent focus of concern, given its association with psychological aspects associated with performance. Researchers commonly explore the intricate relationship between anxiety and performance across various domains, including sports, public speaking, foreign language proficiency, and academic examinations (Cassady, 2010). It is a common expectation that state anxiety tends to emerge before the actual performance and subsequently diminish afterwards. Anxiety has been shown to function as a potent motivational catalyst, augmenting an individual's capacity for heightened focus and improved learning (Öztürk & Uluşahin, 2018), thereby resulting in enhanced performance efficacy, increased efficiency, and heightened intellectual acuity (Guil et al., 2019). However, it is equally acknowledged that an excessive degree of pressure experienced during performance can potentially erode the overall attainment of goals (Gümüş, 2002). Consequently, the multifaceted relationship between anxiety and performance is pivotal in understanding how it can either propel individuals toward success or hinder their accomplishments.

Test anxiety, a form of anxiety, is characterised by a range of emotional, cognitive and behavioural responses that emerge in response to the perception of an examination as a threat to one's future prospects. During the period of examination, the occurrence of physiological signs such as increased heart rate, stomach aches, and sweaty palms are indicative of the emotional aspect of anxiety. Concurrent internal dialogues, including statements such as "I can't do it," "It's too hard," and "If I fail," both before and after exams, reflect the cognitive dimension of anxiety (Doğan-Altun et al., 2017; Onat-Kocabıyık & Donat-Bacıoğlu, 2020).

A body of research has revealed a noteworthy correlation between test anxiety and academic performance. The findings have consistently demonstrated that moderate levels of test anxiety can act as a motivating force, yielding positive outcomes in terms of academic performance. Conversely, when test anxiety escalates to excessive levels, students are prone to exhibit cognitive responses that deviate from anticipated norms, resulting in detrimental effects on their overall performance (Spielberger & Vagg, 1995).

It is well-documented that students experiencing test anxiety often find themselves engulfed in a web of negative thoughts, which can include apprehensions of failure and personal inadequacy in the period leading up to an examination. Moreover, such individuals may encounter impediments in information recall and retrieval during the actual test-taking process. A notable illustration of this relationship emerged in a study involving a cohort of 200 second-year medical school students. The study utilised a multifaceted approach to evaluate test anxiety, employing physiological indicators such as salivary cortisol, heart rate, and blood pressure. The findings of this research established a direct correlation between heightened sympathetic nervous system activity and heightened test anxiety (Hahn et al., 2017).

The primary indicator of test anxiety is the increase in physiological arousal levels. Researchers frequently utilise physiological measurements, such as heart rate, or electrodermal measurements, including skin conductance response, to assess the arousal experienced in examination contexts. These measurements provide an objective perspective through which to assess the extent of students' arousal during the examination process, and researchers have noted that these measurements are challenging to suppress or control (Harley, 2015; Houtveen & de Geus, 2009; Wilhelm & Grossman, 2010).

A wide variety of methodologies can be employed to gather physiological measurements. These include self-reporting, behavioural assessments involving facial expressions, neuroimaging techniques such as functional magnetic resonance imaging (fMRI), and observations of environmental physiological processes, including heart rate and skin conductance (Roos et al., 2021). Among these physiological metrics, heart rate (HR) is of particular significance as a cardiovascular indicator, denoting the frequency of heartbeats within a given time span, commonly expressed as beats per minute (bpm) (Hugdahl, 1995). The normal resting heart rate is known to range between 60 and 100 beats per minute (bpm) (Spodick, 1993; Avram et al., 2018). It has been demonstrated that heart rate provides insights into the autonomic nervous system's modulation of cardiac activity (McCraty et al., 2001). Furthermore, the hypothesis has been postulated that heart rate may serve as a direct reflection of anxious arousal, indicating a positive correlation between higher heart rate and heightened levels of test anxiety (Calvo & Miguel-Tobal, 1998).

It is evident that heart rate, functioning as an objective physiological metric, assumes a pivotal role in the comprehensive exploration of test anxiety. This is evident by drawing a connection between the preceding discourse concerning physiological measurements and heart rate and the subsequent examination of the

interplay between test anxiety and different item types in testing contexts. Beyond this physiological dimension, the impact of the varying types of test items has garnered considerable attention in the research literature.

Researchers have undertaken studies with the objective of scrutinising the intricate relationship between the nature of test items and test anxiety. Significant among these investigations are those that have explored the factors influencing test anxiety, as highlighted by Revilla (2009). Revilla emphasises that test anxiety is influenced by an array of elements, including the specific instructions provided and the formats adopted in the design of the tests themselves. The broad categorization of tests can be divided into objective item types, which include true-false, matching, multiple-choice and open-ended formats. These formats have been the focus of particular scrutiny.

The extant literature presents empirical findings that shed light on the varying levels of test anxiety associated with different item types. For instance, Elfiana (2005) reported that test anxiety is more prevalent when faced with open-ended items as opposed to multiple-choice items. In a similar vein, the research by Oktaviani and Sastrawijaya (2017) revealed that test anxiety is notably higher in response to multiple-choice items as compared to true-false formats. In a similar manner, the work of Wahyuni et al. (2021) has highlighted that multiple-choice items tend to evoke higher levels of test anxiety in comparison to matching item formats.

In parallel with these insights, Aşçı et al. (2015) embarked on an inquiry delving into the effects of the measurement tools used in higher education examination settings. This investigation compared the impact of these tools to an alternative examination format closely linked to real-life situations encountered by the students. The researchers' findings indicated that students demonstrated more favourable responses when confronted with examination items that held personal relevance. This observation was accompanied by a decline in test anxiety and an enhancement in students' self-perceived success.

Furthermore, the research conducted by Temizkan and Sallabaş (2009) dealt with a comparative analysis between written examinations and multiple-choice tests, particularly in the assessment of students' reading comprehension skills. Their results underlined the improved performance of students in the context of multiple-choice items related to reading comprehension.

In addition, Little et al. (2012) explored the differences between written examinations and multiple-choice test formats. They highlighted the benefits of the latter, emphasising the presence of cues that facilitate students' ability to recall the correct answers. Taken together, these diverse investigations provide a nuanced understanding of the complex dynamics linking test anxiety to different item types, and highlight the multifaceted nature of this phenomenon.

Within the existing academic literature, the study of test anxiety typically revolves around the administration of anxiety measurement scales and the subsequent reporting of results based on these scales. However, there is an intriguing gap in the literature that concerns the exploration of the intricate relationship between test anxiety and item types, while measuring this interplay through the lens of physiological indicators, specifically heart rate. Remarkably, no previous instances of such investigations have been documented in the literature. As such, this study represents an innovative and pioneering contribution to the existing scientific discourse.

It is important to acknowledge that a predominant feature of examinations used for student selection and placement in Turkey is the prevalence of multiple-choice item formats. The significance of this research lies in its ability to illuminate the nuanced ways in which the prevalence of multiple-choice items in major academic assessments affects students' test anxiety. The implications of this research extend to the areas of item type selection, not only by national measurement and assessment authorities, but also by educators at different levels of education. Indirectly, students represent a key stakeholder group that may be influenced by the results of this study. The potential implications are far-reaching, including improvements in critical areas such as academic achievement, the cultivation of academic confidence, and the overall well-being of students. These improvements are expected to materialise as test anxiety is reduced through sensible choices of item types in assessments.

In light of these central considerations, the present study aims to investigate the relationship between test anxiety and item type by measuring heart rate variability during different item types (open-ended and multiple-choice) encountered by participants during an exam. The following sub-objectives will be addressed:

- What is the self-perceived level of test anxiety among the participants within the examination setting?
- How do participants' heart rate fluctuations change outside the exam environment?
- What are the variances in participants' anxiety levels as they engage with both multiple-choice and open-ended test items during the examination?

2. Methodology

2.1. Design

This study employs a quantitative research method with a descriptive research design. Descriptive research aims to accurately portray a past or present situation. This design is especially beneficial for offering a comprehensive overview of the variables of interest without altering the study environment (Erkuş, 2013).

2.2. Study Group

Participants in the study were third-year students enrolled in the Department of Guidance and Psychological Counselling. At the beginning of the 2021-2022 academic year, students were given full details of the research project. Potential participants who expressed an interest in volunteering for the study were then identified. Out of the total cohort of 98 students in the class, a total of 76 students agreed to participate. In the following week, these 76 students were asked to complete a demographic information survey. The students' responses to this demographic information survey are tabulated and presented in Table 1.

Table 1. *Distribution of Participating Students by Gender and Age*

Gender and Age	Frequency	Percentage (%)
Female	56	73.7
Male	20	26.3
18-20	35	46.1
21-29	41	53.9

Of the participants, 73.7% were female and 26.3% were male. In terms of age distribution, 46.1% of the students were in the 18-20 year age group, while 53.9% were in the 21-29 year age group. Data were then systematically collected on variables that could potentially influence the students' heart rate. These variables included questions about the presence of chronic and mental health conditions, pregnancy status (for female participants), smoking habits and the predominant hand used by the participants. The tabulated results of this data collection are presented in Table 2.

Table 2. *The Distribution of the Variablest that May Affect the Heart Rate of the Students*

Variables		Frequency	Percentage (%)
Chronic disease	There is	6	7.9
	None	70	92.1
Psychological illness	There is	3	3.9
	None	73	96.1
Smoking	Yes	27	35.5
	No	49	64.5
Dominant hand	Right	64	84.2
	Left	11	14.5
	Both right and left	1	1.3
Pregnancy Status	Yes	0	0
	No	76	100

Within the cohort of participating students, 92.1% reported no chronic conditions, while 7.9% reported the presence of such conditions. Of the students who reported the presence of chronic conditions, specific conditions were mentioned, including one student who reported polycystic kidney disease and another who reported urticaria. The nature of the conditions of the remaining four students was not disclosed.

Conversely, concerning psychological health, a substantial proportion of 96.1% of the students attested to the absence of psychological disorders, while 3.9% acknowledged the presence of such conditions. Of the students who reported having a mental disorder, two specifically reported experiencing anxiety, while one student reported having an impulse control disorder.

Regarding smoking habits, 35.5% of the students were smokers and 64.5% were non-smokers. Of the students, 84.2% predominantly used their right hand, 14.5% predominantly used their left hand and 1.3% used both hands equally. All responses regarding pregnancy status indicated that none of the students were pregnant. On the basis of these variables with potential effects on heart rate, the inclusion criteria for the study were structured to include students who did not report any chronic or mental health conditions, were identified as non-smokers, and preferred to use their right hand.

After careful consideration of the above criteria, the selected students were assessed using the Test Anxiety Inventory (TAI) and the State-Trait Anxiety Inventory (STAI). On the basis of the scores derived from these instruments, a subgroup of 32 students was identified as meeting the criteria for inclusion in the working group, with an average level of anxiety. As a result, a total of 20 students remained consistently engaged in all research-related procedures throughout the academic year, thus forming the cohesive working group for this study.

2.3. Data Collection Tools and Procedure

Demographic Information Form: A demographic information form was administered to collect comprehensive details of the students' demographic characteristics. This included questions on gender, age, medical conditions, mental disorders, smoking habits, coffee consumption and pregnancy status. The formulation of this form was inspired by "The Health Survey-SF-36" and the research conducted by Hardacre Cerqueira (2015).

Test Anxiety Scale: The Test Anxiety Scale, developed by Benson and El-Zahhar (1994) and adapted into Turkish by Akın et al. (2012), was used to determine students' test anxiety levels. Benson and El-Zahhar (1994) provided the final version of the scale, consisting of 20 items and 4 subscales, in both American and Egyptian samples. The Cronbach's alpha reliability coefficient of the scale varied between .76 and .91 in the American sample. In the Egyptian sample, the Cronbach's alpha reliability coefficients were .60 for the tension subscale, .68 for the thoughts unrelated to exams subscale, .84 for the worry subscale, and .73 for the physical symptoms subscale (Benson & El-Zahhar, 1994). In the adapted Turkish version of the scale by Akın et al. (2012), it was found that the twenty items were retained in their original form, forming four dimensions, and the four-dimensional model showed a good fit (χ^2 = 332.20, sd = 160, RMSEA = .056, CFI = .92, IFI = .92, GFI = .91, SRMR = .051). The internal consistency coefficients of the scale were calculated as .78 for the tension subscale, .77 for the physical symptoms subscale, .71 for the worry subscale, .80 for the thoughts unrelated to exams subscale and .88 for the total scale. The test-retest coefficients obtained at three-week intervals were .75, .64, .74 and .70 for the subscales and .65 for the total scale. The item-total correlation coefficients of the scale ranged from .37 to .58. The Cronbach's alpha of the test anxiety scale used in this study was found to be .89.

State-Trait Anxiety Inventory: The State-Trait Anxiety Inventory, developed by Spielberger (1966), is a 40-item scale that reflects state and trait anxiety levels. The adaptation of the inventory to Turkish culture was carried out by Öner and Le Compte (1983). The reliability values for the trait anxiety scale ranged from .83 to .87 and for the state anxiety scale from .94 to .96 (Cronbach's alpha). Within the framework of this study, the reliability values of the data obtained were found to be between .80 and .86 for the trait anxiety scale and between .91 and .93 for the state anxiety scale (Cronbach's alpha).

The Polar M430 Wristband: The Polar M430 wristband uses optical technology to measure heart rate through a sensor placed on the back of the wristband (Polar, 2018). In the research, the Polar M430 device was used to measure the heart rates of the working group both outside of the examination environment and during the examination. The Polar M430 has emerged as a prominent data collection tool in various studies (Batalik et al., 2018; Haney, 2018; Kons et al., 2018; Kons et al., 2019), with a parallel focus on assessing its validity and reliability (Hanzel et al., 2018; Patterson et al., 2018; Shryack et al., 2018).

For data collection in the study, a total of four tests consisting of open-ended and multiple-choice items were developed to measure the gains from the Principles and Techniques of Psychological Counselling course in the preliminary application phase of the study. A content validity table and expert opinion were used in the development of these tests. Two experts were involved in this process - one from the Department of Measurement and Evaluation in Education and the other from the Department of Guidance and Psychological Counselling. The experts provided recommendations, particularly for the formulation of open-ended items.

These open-ended items were designed to create real-life situations related to the subject to be measured and required students to reflect and make inferences related to the subject. In addition, two multiple-choice tests were developed, each consisting of 20 items, and the data obtained from these tests were examined for reliability using the KR-20 method of internal consistency. The reliability of the data was found to be .82 and .79 respectively. The reliability of the open-ended tests was assessed by having a faculty member from the field of guidance and psychological counselling read the tests at different times and calculating the coefficient of agreement between the scores given at different intervals. The coefficients of agreement obtained were .91 and .93 respectively.

The main application of the study was during the counselling skills course. This course covered all the gains of the course and included a 30-item multiple-choice test and a 5-item open-ended test. The reliability values of the data obtained from these tests were found to be .84 and .90 respectively.

The reason for choosing these two courses for the preliminary and main applications of the study is that they are interrelated courses. The practical aspect of the Principles and Techniques of Psychological Counselling course is covered in the Counselling Skills course. In addition, both courses are foundational courses in the curriculum of the Department of Guidance and Psychological Counselling. The success of students in these courses is of paramount importance to their ability to meet the demands of the profession in their future careers. In this context, the desire to succeed in these two courses may lead students to engage in certain emotional behaviours. According to Bloom et al. (1956), students' entry behaviour in the affective domain is associated with 25% of cognitive success. From this perspective, the fear of failure also leads to anxiety in this course.

In order to determine the resting and writing heart rates of the working group outside of the examination environment, a subgroup of third year students from the Department of Guidance and Psychological Counselling was selected. These students were initially measured for their resting and writing heart rates in the Principles and Techniques of Psychological Counselling course. This initial measurement was to establish their baseline heart rate. The selected students were then taken to a classroom within the faculty and asked to remain in a stationary position for 10 minutes. At the end of the 10 minutes, heart rate monitors were attached to their wrists and measurements were taken for five minutes. Students were then given a paragraph of text and asked to copy it onto another sheet of paper within five minutes.

In addition, the students were informed at the beginning of the data collection process not to consume coffee before the measurement sessions. At the end of the study, students who completed all applications were given an external storage device as a token of appreciation for their support. The students were informed of this gift at the beginning of the study.

2.4. Data Analysis

The Test Anxiety Scale and the State-Trait Anxiety Inventory for State Anxiety (STAI-S) were used to determine the level of test anxiety perceived by the working group in the test environment. The non-parametric Friedman test was used to assess the significance of the differences between the data obtained from these instruments. Variations in heart rate for both open-ended and multiple-choice test types are presented in the figures, and evaluations were made based on total variances.

2.5. Ethical

All procedures performed in our study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the Helsinki Declaration of 1964 and its subsequent amendments or comparable ethical standards. The Rectorate of Trakya University, Chair of the Ethics Committee for Social and Human Sciences Research approved the research protocol under the application number 08/05.

3. Findings

3.1. What is the self-perceived level of test anxiety among the participants within the examination setting?

To determine the level of test anxiety perceived by the working group in the examination environment, the significance of the differences between the rank means of the scores obtained from the Test Anxiety Scale and

the State-Trait Test Anxiety Inventory of the 20 students in the working group was tested. The non-parametric Friedman test was used for this purpose. The results of the test are presented in Table 3.

Table 3. *Friedman Test's Results*

	n	Mean rank	χ2	df	p
State test anxiety	20	1.78	13.241	2	.001
Trait test anxiety	20	2.65			
Test anxiety	20	1.58			

The non-parametric Friedman test was used to test the significance of the differences in the mean ranks of the state exam anxiety, trait exam anxiety and test anxiety scores. The results showed a significant difference between the mean ranks of all three scores ($\chi 2 = 13.241$; $p<.01$). To determine which scores showed significant differences, Wilcoxon tests were performed. The results indicated that there was a significant difference between state test anxiety and trait test anxiety ($Z = 2.431$; $p<.05$) and between test anxiety and trait test anxiety ($Z = 3.941$; $p<.001$). Based on these results, the mean ranks of the scores obtained from the Trait Test Anxiety Inventory and the Test Anxiety Scale were higher than the mean ranks of the scores obtained from the State Test Anxiety Inventory. Consequently, the students included in the study had higher levels of trait test anxiety.

3.2. How do participants' heart rate fluctuations change outside the exam environment?

For the second objective of the study, the smartwatch (Polar M430) was used to measure students' heart rates during resting and writing situations to assess their heart rates outside of the test environment. The smartwatch recorded the heart rate data for each second throughout the measurement period. At this point, the changes in students' heart rates were considered as indicators of anxiety. Therefore, to illustrate the changes in heart rates during resting and writing, the variance values of heart rates are presented in Figure 1 below.

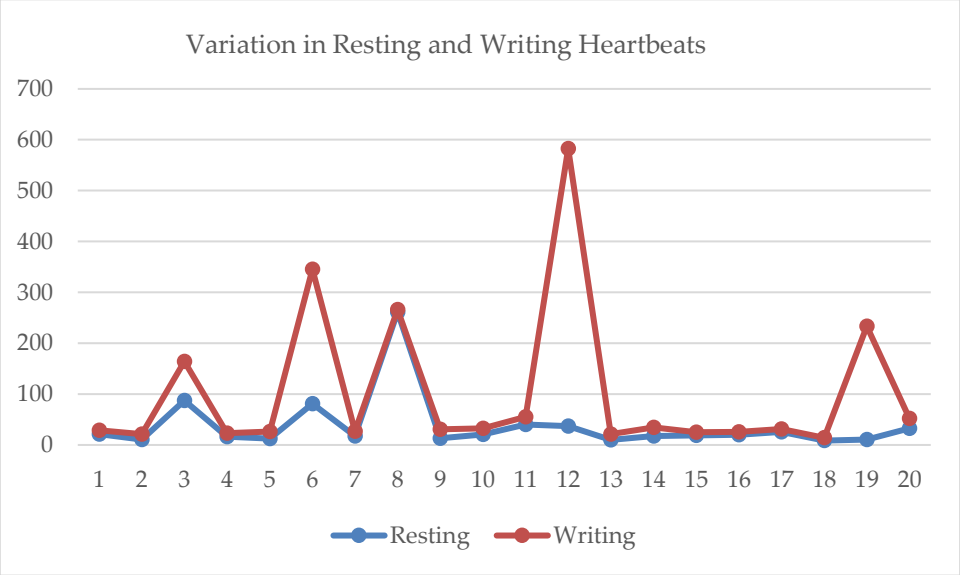

Figure 1. *Variation in Heart Rate at Rest and Writing*

Students' heart rates were measured for the first time in resting and writing situations. By 'resting' we refer to the students' steady state 10 minutes after entering the classroom, and by 'writing' we refer to the verbatim reproduction of an informative paragraph about education. The horizontal axis represents the students, while the vertical axis shows the variance in their heart rate over a 10-minute period. As a result, their heart rates were higher than expected because it was their first experience of this measurement process. In order to understand the reason for this, interviews were conducted with the students and when examining the graph, the 12th student, who had the highest change in heart rate during writing, expressed that it was their first time experiencing such a measurement and they were unsure of what to expect during the process, which could have led to a higher than expected heart rate. In particular, they mentioned feeling anxious, as if they were taking an exam, when they saw the paper during the writing task.

3.3. What are the variances in participants' anxiety levels as they engage with both multiple-choice and open-ended test items during the examination?

For the third objective of the study, the initial application was conducted to determine the anxiety levels of the working group during the exams consisting of multiple-choice and open-ended items.

For this purpose, a total of four test sessions were administered, including pre-midterm and pre-final exams during the autumn semester of the Principles and Techniques of Psychological Counselling course. These tests included both multiple-choice and open-ended tests. The variance values of the students' heart rates obtained during the tests are presented in Figure 2 below.

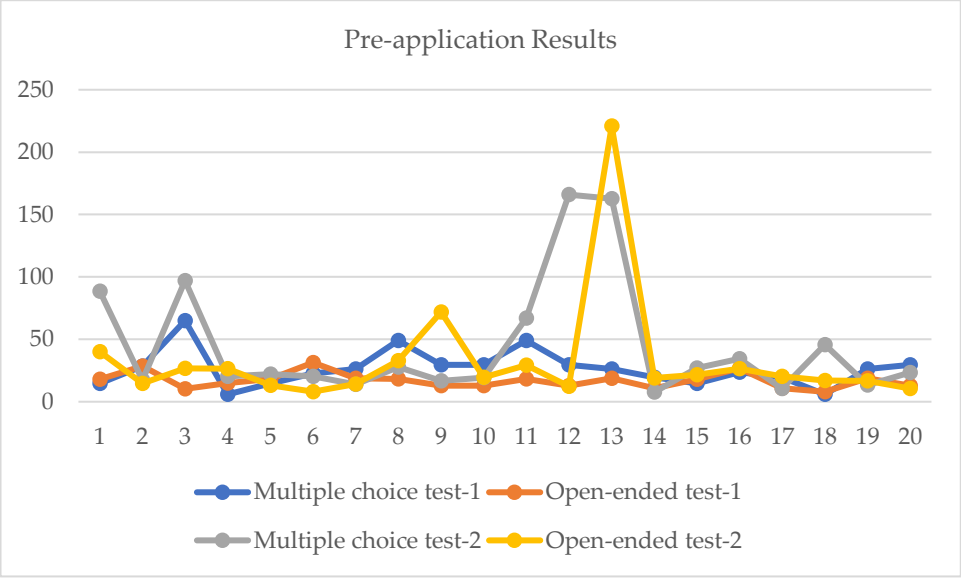

Figure 2. *Variation in Pre-Application Results*

Figure 2 shows the variations in the students' heart rates during the first application of the four tests. According to the results, the 12th and 13th graders show higher variations in heart rate compared to the other students, especially during the open-ended test. In general, heart rate variability is greater on multiple-choice tests than on open-ended tests. In order to better assess the overall changes in heart rate for both item types, Table 4 shows the sum of students' heart rate variations over the four test sessions.

Table 4. *Total Variances of Pre-Treatment Results*

	Total variances
Multiple choice test 1	529.0998
Open-ended test 1	339.7596
Multiple choice test 2	901.1302
Open-ended test 2	662.0533

Looking at the total variances of the heart rates obtained in the first applications in Table 4, it is evident that the total variance of the heart rates in the first multiple-choice test administered is higher than the total variance of the heart rates in the open-ended test. Similarly, the total variance of heart rates in the second multiple-choice test is also higher than the total variance of heart rates in the open-ended test. Therefore, in both cases, the variations in heart rate on the multiple-choice tests were greater than those on the open-ended test. Consequently, the students experienced more test anxiety in the multiple-choice tests during the first applications. The comparisons in this table are made between 'Multiple-choice test 1' and 'Open-ended test 1', and between 'Multiple-choice test 2' and 'Open-ended test 2', as these tests measure the same outcomes. Multiple-choice test 1 and free-response test 1 were administered two weeks before the mid-term exam, while multiple-choice test 2 and free-response test 2 were administered two weeks before the final exam.

In the main application of the study, during the spring semester, multiple-choice and open-ended tests were administered separately for two consecutive weeks before the final exam of the Counselling Skills course. The changes in heart rate obtained from these two tests are shown in Figure 3.

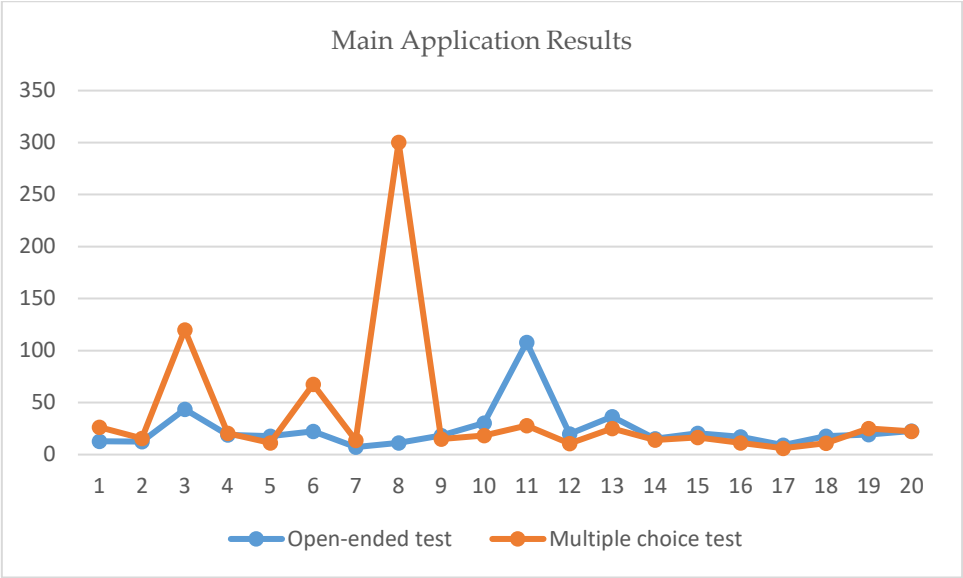

Figure 3. *Variation in Main Application Results*

Looking at Figure 3, it is noticeable that there is a significant difference in the change in heart rate between the multiple-choice and free-response tests, which is particularly evident in student 8. Similar patterns of difference are also observed for the 3rd, 6th and 11th students. However, in the case of the 11th student, the change in heart rate during the open-ended test was found to be higher. In order to better assess the overall changes, Table 5 also shows the total variances of the students' heart rates over the two test administrations.

Table 5. *Total Variance of Main Application Results*

	Total variance
Multiple choice test	778,6354
Open-ended test	480,5703

Table 5 shows that the total variance of heart rates obtained from the multiple-choice test is higher than the total variance obtained from the free-response test. Accordingly, students showed a greater change in heart rate during the multiple-choice test than during the open-ended test. Similar results were obtained in both the preliminary and main applications of the study. In both cases, the multiple-choice tests produced more significant changes in students' heart rates.

4. Discussion and Conclusion

In this study, the use of heart rate, a physiological indicator, to assess students' test anxiety provides a direct measure that goes beyond psychological measures of anxiety. Psychological tests often used in test anxiety research are developed on the basis of subjective experience and may not directly reflect behaviour at the time of testing (Tellegen, 1985). Although these tests are useful for a standard assessment of behaviour, heart rate, which is a more specific and physiological response, more clearly reflects tension at the time of the test. It has been shown in the literature that heart rate is related to emotional states such as stress, anxiety and depression (Clarke et al., 2014; Shinba et al., 2008). Therefore, in order to understand a complex phenomenon such as test anxiety, the evaluation of physiological data also provides a more comprehensive approach to this phenomenon. This study highlights the potential of heart rate in test anxiety research and offers a new perspective for future studies.

This study found that students' heart rate variability was higher during multiple-choice tests than during open-ended tests. This finding suggests that multiple-choice tests may induce higher levels of anxiety in students. On the other hand, open-ended tests may cause less anxiety because they allow students to use more flexible thinking skills.

This finding highlights the importance of item type as a factor influencing students' anxiety levels. In this regard, Önder (2008) investigated the relationship between test anxiety and test type. The study found significant evidence that open-ended tests cause less anxiety among students. One possible reason why multiple-choice tests elicit higher levels of student anxiety is that they are generally perceived as less reliable

in assessing students' higher cognitive levels. They are often criticised for measuring recognition skills rather than application, analysis or evaluation skills (Campbell et al., 2002). This perception may lead students to believe that they are being unfairly assessed and make them feel more anxious. However, Tekin (2000) argues that well constructed multiple-choice items can assess not only knowledge but also comprehension, application, analysis, synthesis and even evaluation behaviour. Nevertheless, Tekin (2000) also points out that multiple-choice items may not adequately measure creativity and may not contribute to the development of creativity.

Another reason why multiple-choice items may cause anxiety in students is their use in high-stakes examinations. In Turkey, all exams for the transition of students to the next educational level, such as high school and university entrance exams administered by the Ministry of National Education and the Student Selection and Placement Center, consist entirely of multiple-choice items (Önder, 2008). Similarly, multiple-choice tests are widely used as the primary assessment method in many education systems worldwide. This is due to their perceived efficiency and cost-effectiveness in assessing student learning. However, as mentioned above, relying solely on multiple-choice tests as the primary method of assessment can lead to higher levels of anxiety in students, which can negatively affect their learning and academic performance (Crocker & Schmitt, 1987). This situation creates pressure on students to perform well in these examinations, leading to increased anxiety and stress. Crocker and Schmitt (1987) examined the performance of students with different levels of test anxiety on multiple-choice items. Their study concluded that individuals with low test anxiety performed better on multiple-choice items, while those with high test anxiety performed poorly. However, Naveh-Benjamin et al (1981) found that individuals with high test anxiety generally performed better on multiple-choice items that required less active recall. Conversely, Birenbaum and Feldman (1995) showed that students with high test anxiety had more negative attitudes towards open-ended test items than those with low test anxiety.

Multiple-choice items typically focus on either the correct or incorrect answer. This limitation is referred to in the literature as the need for full competence to receive a score (Wongwiwatthananukit et al., 2000). This also indicates that partial knowledge is not taken into account (Diedenhofen & Musch, 2015). Consequently, an individual who knows that they will not receive any points if they cannot answer a multiple-choice item correctly may perceive this situation as test anxiety.

Open-ended tests, on the other hand, may cause less anxiety for students because they allow for more flexible thinking. In addition, open-ended items are seen as a more authentic assessment of student learning. They allow students to demonstrate their understanding of a topic in their own words, which can be seen as a more accurate reflection of their knowledge. In addition, open-ended items allow for more extensive use of thinking skills and provide students with the opportunity to find an answer they are confident in, thus facilitating lower anxiety levels (Tekin, 2000; Tekindal, 2002). As a result, students may feel that the assessment process is more flexible and therefore experience lower levels of anxiety.

According to McKeachie (1986), open-ended items, which require longer responses, are valuable tools for students to organise, synthesise and interpret information in their own words. Unlike multiple-choice items, which require test-takers to select the correct answer from a set of options, open-ended items require test-takers to construct their responses (Kastner & Stangla, 2011; Messick, 1994; Rodriquez, 2002; Roid & Haladyna, 1982). As such, open-ended items can provide a more detailed and in-depth assessment of an individual's knowledge (Rodriquez, 2002).

In conclusion, although multiple-choice tests are still widely used in our national education system, it is crucial to be aware of their limitations and their potential negative impact on students' anxiety levels. The Ministry of National Education should actively work towards implementing new assessment methods, such as open-ended items and authentic assessments, which can provide a more comprehensive assessment of student learning and reduce anxiety levels. While acknowledging that multiple-choice tests have their advantages in terms of efficiency, objectivity and the ability to assess a large number of students quickly, it is important to recognise the limitations of this method.

In fact, on 9 September 2023, the Ministry of National Education took a historic decision and decided to hold common exams consisting of open-ended items in schools (except for the primary level) (Official Gazette of the Republic of Turkiye, 2023). For the 2023-2024 school year, the Ministry of National Education stated that

multiple-choice items will be used in the common written exams to be held in all provinces and districts, but open-ended items may also be used in these exams after the necessary infrastructure is prepared. Otherwise, questions in all common written examinations and all other written examinations will consist of open-ended or short-answer questions. Questions for the common written exams to be organised by the Ministry will be prepared by the General Directorate of Measurement, Evaluation and Examination Services and sent to the schools. Teachers working at the school will administer the exam and the school or the Directorate of Measurement and Evaluation Centre will evaluate it. This is a new and important decision for the Turkish examination system, where multiple-choice items dominate. However, when open-ended items are used in exams, it is more desirable not to make them compulsory, but to diversify the types of items according to the information that is to be obtained from the student.

It is important to remember that both item types have their advantages and disadvantages. Given the evidence that multiple-choice items are less conducive to creative thinking and writing (Haladyna, 1997) and that open-ended items may raise concerns about the objectivity of scoring (Brookhart, 1994), it is important not to overlook the potential anxiety that students may experience in both cases. It is therefore important to provide students with the necessary tools and strategies to manage test anxiety. Universities can offer workshops and seminars on test-taking strategies, relaxation techniques and time management to help students cope with exam stress.

Given the finding that multiple-choice tests may induce higher levels of anxiety in students, it may also be beneficial to limit their use in university settings. However, it is important to remember that multiple-choice tests also have their advantages, such as efficiency, objectivity and the ability to assess a large group of students quickly.

5. Recommendations

A balance should be struck between the two types of items, recognising that no single type of item can adequately measure all cognitive processes and learning outcomes. The use of different item types allows for a wide range of measurements. In particular, on an international scale, large-scale applications (e.g. Programme for International Student Assessment, Trends in International Mathematics and Science Study, GRE (Graduate Record Examinations)) include both multiple-choice and open-ended items (DeCarlo et al., 2011; Mariano, 2002). Such assessments, which allow students to demonstrate their understanding in their own words, can provide a more realistic assessment of student learning. In this way, we can address students' test anxiety and increase the effectiveness of educational assessment.

In future research, a similar study can be repeated with more participants and the findings can be compared by measuring heart rate in exams of more than one course.

6. Limitations

The number of participants can be considered within the scope of the limitations of the research. This study is a product of the TÜBAP project numbered 2021/05. Within the scope of the project, only 15 watches were budgeted. Therefore, the study group was limited to data collected from 30 participants in two different sessions across two separate groups. A series of measurements were conducted as part of the study, and 20 individuals who participated in all measurements constituted the study group. This is a limitation of the research. Similar studies can be conducted with a larger number of participants by increasing the number of watches through projects with higher budgets.

Disclosure statement: This paper was edited and refined with the assistance of Google's Gemini (version as of June 2024), which complements the human editorial process. The human authors critically assessed and validated the content to maintain academic rigor. They also evaluated and addressed potential biases inherent in AI-generated content. The final version of the paper is solely the responsibility of the human authors.

7. References

Akın, A., Demirci, İ., & Arslan, S. (2012). Revised Test Anxiety Scale: Validity and reliability study. *Journal of Educational Sciences & Practices, 11*(21), 103-118.

Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Berger, A. C., & Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science, 213*(4504), 220–222. https://doi.org/10.1126/science.6166045

Aşçı, M., Işlak, O., & Güler, Ş. (2015). The effects of evaluation instruments used for assessment and measurement at higher education on students' exam anxiety the example of the faculty of education at Celal Bayar University. *Celal Bayar University Journal of Social Sciences, 13*(4), 185-197. http://dx.doi.org/10.18026/cbusos.51090

Avram, R., Kuhar, P., Vittinghoff, E., Aschbacher, K., Tison, G., Pletcher, M., et al. (2018). Redefining normal resting heart rate values using Big Data. *Circulation, 138*(Suppl_1), A15098. https://doi.org/10.1161/circ.138.suppl_1.15098.

Batalik, L., Dosbaba, F., Hartman, M., Batalikova, K., & Spinar, J. (2020). Benefits and effectiveness of using a wrist heart rate monitor as a telerehabilitation device in cardiac patients: A randomized controlled trial. *Medicine, 99*(11), e19556. https://doi.org/10.1097/MD.0000000000019556

Benson, J., & El-Zahhar, N. (1994). Further refinement and validation of the Revised Test Anxiety Scale. *Structural Equation Modeling: A Multidisciplinary Journal, 1*(3), 203-221. https://doi.org/10.1080/10705519409539975

Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research, 40*(1), 90-98. https://doi.org/10.1080/0013188980400109

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain* (pp. 201-207). McKay.

Brookhart, S. M. (1994). Teachers' grading practice and theory. *Applied Measurement in Education, 7*(4), 279-301. https://doi.org/10.1207/s15324818ame0704_2

Calvo, M. G., & Miguel-Tobal, J. J. (1998). The anxiety response: Concordance among components. *Motivation and Emotion, 22*(3), 211–230. https://doi.org/10.1023/A:1022384022641

Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. In *Annual meeting of the mid-western educational research association*, Columbus, OH (pp. 17-18).

Cassady, J. C. (2010). *Anxiety in schools: The causes, consequences, and solutions for academic anxieties.* Peter Lang Publishing.

Cattell, R. B., & Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety.* Ronald.

Clarke, S., Horeczko, T., Cotton, D.,& Bair, A. (2014). Heart rate, anxiety and performance of residents during a simulated critical clinical encounter: a pilot study. *BMC Med Educ 14*(153). https://doi.org/10.1186/1472-6920-14-153

Crocker, L., & Schmitt, A. (1987). Improving multiple-choice test performance for examinees with different levels of test anxiety. *The Journal of Experimental Education, 55*(4), 201-205. https://doi.org/10.1080/00220973.1987.10806454

DeCarlo, L. T., Kim, Y., & Johnson, M. S. (2011). A hierarchical rater model for constructed responses, with a signal detection rater model. *Journal of Educational Measurement, 48*(3), 333-356. https://doi.org/10.1111/j.1745-3984.2011.00143.x

Diedenhofen, B., & Musch, J. (2015). Empirical option weights improve the validity of a multiple-choice knowledge test. *European Journal of Psychological Assessment, 33*(5), 336–344, 2015. https://doi.org/10.1027/1015-5759/a000295.

Doğan-Altun, Z., Avcı, G., Özüpekçe, S., & Gümüş, N. (2017). College of education students' public personnel selection exam anxiety. *Turkish Studies, 12*(33), 151-166. http://dx.doi.org/10.7827/TurkishStudies.12743

Elfiana, S. (2005). *Perbedaan kecemasan tes antara tes berbentuk pilihan ganda dan tes berbentuk esai.* Jakarta: Fakultas Psikologi, Universitas Katolik Atma Jaya Jakarta.

Erkuş, A. (2013). *Davranış bilimleri için bilimsel araştırma süreci [Scientific research process for behavioral sciences]*. Seçkin Publishing.

Guil, R., Gómez-Molinero, R., Merchan-Clavellino, A., Gil-Olarte, P., & Zayas, A. (2019). Facing anxiety, growing up. Trait emotional intelligence as a mediator of the relationship between self-esteem and university anxiety. *Frontiers in Psychology*, 10, 567. https://doi.org/10.3389/fpsyg.2019.00567

Gümüş, M. (2002). *Profesyonel futbol takımlarında puan sıralamasına göre durumluk kaygı düzeylerinin incelenmesi [Examination of state anxiety levels in professional football teams according to score rankings]* [Master thesis]. Sakarya University, Turkiye.

Hahn, H., Kropp, P., Kirschstein, T., Rücker, G., & Müller-Hilke, B. (2017). Test anxiety in medical school is unrelated to academic performance but correlates with an effort/reward imbalance. *PloS One*, *12*(2): e0171220. https://doi.org/10.1371/journal.pone.0171220

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Allyn & Bacon.

Haney, A. C. (2018). *Young female college millennials' intent for behavior change with wearable fitness technology* [Doctoral dissertation]. Walden University, United States.

Hanzel, O., Shryack, G., Patterson, J., Willoughby, C.A., & Smith, B.W. (2018). Comparison of reported energy expenditure from Polar M430 vs. Polar V800 vs. Actual energy expenditure. *Medicine & Science in Sports & Exercise, 50*(5S), 661. https://doi.org/10.1249/01.mss.0000538182.95046.49

Hardacre Cerqueira, B. (2015). *Behavioral and psychophysiological correlates of conversation* [Doctoral dissertation]. UCLA.

Harley, J. M. (2015). Measuring emotions: A survey of cutting-edge methodologies used in computer-based learning environment research. In S. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 89–114). Academic Press, Elsevier.

Houtveen, J. H., & de Geus, E. J.C. (2009). Noninvasive psychophysiological ambulatory recordings: Study design and data analysis strategies. *European Psychologist, 14*(2), 132–141. https://doi.org/10.1027/1016-9040.14.2.132

Hugdahl, K. (1995). *Psychophysiology: The mind–body perspective.* Harvard University Press.

Kapur, S., Debar V., & Kaur R. (2019). Lifestyle affects the pre-competitive state anxiety: Study on male and female basketball players. *International Journal of Physiology, Nutrition and Physical Education, 4*(1), 1079-1082.

Karasar, N. (2006). *Bilimsel araştırma yöntemi [Scientific research method]* (16th Press). Nobel.

Kastner, M., & Stangla, B. (2011). Multiple choice and constructed response tests: Do test format and scoring matter?. *Procedia-Social and Behavioral Sciences, 12*, 263-273. https://doi.org/10.1016/j.sbspro.2011.02.03

Kons, R. L., Detanico, D., Ache-Dias, J., & Dal Pupo, J. (2019). Relationship between physical fitness and match-derived performance in judo athletes according to weight category. *Sport Sciences for Health, 15*, 361-368. https://doi.org/10.1007/s11332-018-00524-y

Kons, R. L., Franchini, E., & Detanico, D. (2018). Relationship between physical fitness, attacks and effectiveness in short- and long-duration judo matches. *International Journal of Performance Analysis in Sport, 18*, 1024 - 1036. https://doi.org/10.1080/24748668.2018.1545198

Lazarus R. S. & Folkman S. (1984). Coping and adaptation. In W. D. Gentry (Ed.), *the handbook of behavioral medicine* (pp. 282–325). Guilford.

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science, 23*(11), 1337–1344. https://doi.org/10.1177/0956797612443370

Mariano, L. T. (2002). *Information accumulation, model selection and rater behavior in constructed response student assessments* [Doctoral dissertation]. Carnegie Mellon University.

McCraty, R., Atkinson, M., Tomasino, D., & Stuppy, W. P. (2001). Analysis of twenty-four hour heart rate variability in patients with panic disorder. *Biological Psychology, 56*(2-3), 131–150.

McKeachie, W. J. (1986). *Teaching tips: A guidebook for the beginning college teacher.* D.C. Heath.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23. https://doi.org/10.3102/0013189X023002013

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y.-g., & Holinger, D. P. (1981). Test anxiety: Deficits in information processing. *Journal of Educational Psychology, 73*(6), 816–824. https://doi.org/10.1037/0022-0663.73.6.816

Official Gazette of the Republic of Turkiye (2023). *Ministry of national education measurement and evaluation regulation.* Access address: https://www.resmigazete.gov.tr/eskiler/2023/09/20230908-2.htm (In Turkish).

Oktaviani, K. S. & Sastrawijaya, Y. (2017). Bentuk tes objektif dan kecemasan pada pembelajaran membaca huruf hiragana bahasa jepang. *Jurnal Evaluasi Pendidikan, 8*(1), 50-56. https://doi.org/10.21009/JEP.081.07

Onat-Kocabıyık, O., & Donat-Bacıoğlu, S. (2020). The role of the rumination and gender on prediciton state test anxiety. *Kastamonu Education Journal, 28*(5), 2165-2172. https://doi.org/10.24106/kefdergi.4087

Önder, O. (2008). *Influence of exam preparation with multiple choice and essay type questions on mathematics achievement and test anxiety levels* [Master thesis]. Marmara University, Turkiye.

Öner, N., & Le Compte, A. (1983) *Durumluk/sürekli kaygı envanteri el kitabı [State/trait anxiety inventory manual].* Boğaziçi University Publications.

Öztürk, M. O., & Uluşahin, N. A. (2018). *Ruh sağlığı ve bozuklukları [Mental health and disorders]* (18th Press). Nobel Tıp Kitabevi.

Patterson, J., Hanzel, O., Shryack, G., Willoughby, C., & Smith, B. (2018). Comparison of two heart rate technologies to predict VO2Max. *Medicine & Science in Sports & Exercise, 50*(5S), 674.

Polar (2018). *Polar U.S.A. products.* Access address: https://www.Polar.Com/Usen/Products

Revilla, L. (2009). Kecemasan menghadapi tes (test anxiety) dan dampaknya terhadap aktivitas belajar. *Dinamika Ilmu, 9*(2).

Rodriquez, M. C. (2002). Choosing an item format. In Tindal, G. & Haladyna, TM (Ed.). *Large-Scale assessment programs for all students* (pp. 213-231). Lawrence Erlbaum Associates Publishers.

Roid, G. H., & Haladyna, T. M. (1982). *A technology for test-item writing.* Academic Press.

Roos, A. L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review, 33*(2), 579-618. https://doi.org/10.1007/S10648-020-09543-Z

Şahin, M. (2019). Korku, kaygı ve kaygı (anksiyete) bozuklukları [Fear, worry and anxiety disorders]. *Eurasian Journal of Social and Economic Research, 6*(10), 117-135.

Shinba, T., Kariya, N., Matsui, Y., Ozawa, N., Matsuda, Y., & Yamamoto, K. I. (2008). Decrease in heart rate variability response to task is related to anxiety and depressiveness in normal subjects. *Psychiatry and Clinical Neurosciences, 62*(5), 603-609. https://doi.org/10.1111/j.1440-1819.2008.01855.x

Shryack, G., Patterson, J., Hanzel, O., Willoughby, C., & Smith, B. (2018). Accuracy of The Polar M430 to Predict Vo2 using optical technology. *Medicine and Science in Sports and Exercise, 50*(5S), 674.

Spiegler, M. D., Morris, L. W., & Liebert, R. M. (1968). Cognitive and emotional components of test anxiety: Temporal factors. *Psychological Reports, 22*(2), 451–456. https://doi.org/10.2466/pr0.1968.22.2.451

Spielberger C. D. (1966). Theory and research on anxiety. In Spielberger C. D. (Ed.), *Anxiety and behavior,* (pp. 413-428). Academic Press.

Spielberger, C. D. (1980). *Test Anxiety Inventory: Preliminary professional manual.* Consulting Psychologists Press.

Spielberger, C. D., & Vagg, P. R. (Eds.). (1995). *Test anxiety: Theory, assessment, and treatment.* Taylor & Francis.

Spodick, D. H. (1993). Is there such a thing as normal sinus rate? *British Heart Journal, 70*(5), 486-487. https://doi.org/10.1136/hrt.70.5.486-a

Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]* (14th Press). Yargı Publishing.

Tekindal, S. (2002). *Okullarda ölçme ve değerlendirme yöntemleri [Measurement and evaluation methods in schools].* Evrim Publishing.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Lawrence Erlbaum Associates, Inc.

Temizkan, M., & Sallabaş, M. E. (2009). Compare the attitudes of teacher candidates related to reading and writing activities. *Electronic Journal of Social Sciences, 8*(27), 155-176.

Wahyuni, L. D., Citraini, R., Hutomo, B. A., & Rakhman, G. G. F. (2021). Anxiety and test form: The differences of test anxiety levels in terms of test form. *JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia), 10*(2), 184-192.

Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biol Psychol, 84*(3), 552–569. https://doi.org/10.1016/j.biopsycho.2010.01.017

Wongwiwatthananukit, S., Popovich, N. G., & Bennett, D. E. (2000). Assessing pharmacy student knowledge on multiple-choice examinations using partial-credit scoring of combined response multiple-choice items. *American Journal of Pharmaceutical Education, 64*(1), 1-10.