

Examination of Scales for Test Anxiety with Thematic Content Analysis

Nurevsa Münevver Bahçeci^a  Yıldız Yıldırım^b  Melek Gülşah Şahin^c 

^a Psychologist, Artı Enstitü, Ankara, Türkiye, nmbahceci@gmail.com

^b Assist. Prof. Dr., Aydın Adnan Menderes University, Aydın, Türkiye, yildizyldrm@gmail.com

^c Prof. Dr., Gazi University, Ankara, Türkiye, mgulsahsahin@gazi.edu.tr

ABSTRACT

This study aims to examine the descriptive and psychometric properties of test anxiety scales in Türkiye through thematic content analysis. A total of 12 scales were obtained by means of the review of studies conducted since 2013. Two studies were not evaluated because they failed to meet the requisite criteria. Document analysis was used as the data collection method. Many of the scales examined consisted of articles. The majority of scale development and adaptation studies were conducted between 2016 and 2018. Most of the scales were of the Likert-type and the number of items ranged from 11 to 30. The multidimensional structure of test anxiety was not considered in some scales, while a number of scales were developed to measure the unidimensional structure of test anxiety. In reliability estimates, the Cronbach alpha coefficient was calculated for all scales, but different reliability estimation methods were also used, including Composite reliability, Spearman-Brown Split half reliability, and test-retest. Cronbach's alpha coefficient was calculated as a minimum of 0.70 and a maximum of 0.94. When the validity data of the scales were reviewed, it was observed that construct validity was the most frequently examined construct, followed by criterion validity. However, content validity was not investigated in any of the studies. Findings regarding EFA and CFA values in the studies on construct validity and the scales used in the studies on criterion validity are presented. At the same time, the analysis of the psychometric properties revealed certain deficiencies, and suggestions were made accordingly.

Article Type
Research

Article Background
Received:
02.08.2024
Accepted:
03.11.2024

Keywords
Test Anxiety, Scale
Development, Scale
Adaptation, Thematic
Content Analysis,
Psychometric Properties

To cite this article: Bahçeci, N. M., Yıldırım, Y., & Şahin, M. G. (2024). Examination of scales for test anxiety with thematic content analysis. *International Journal of Turkish Education Sciences*, 12 (3), 1513-1548. <https://doi.org/10.46778/goputeb.1526461>

Corresponding Author: Yıldız Yıldırım, e-mail: yildizyldrm@gmail.com

Introduction

Test anxiety is a common issue affecting students across different educational levels from primary school to graduate education (Chapell et al., 2005). People who suffer from test anxiety involve feelings of unease, apprehension, and fear that individuals experience before, during, or after examinations without noticing age. While the intensity of anxiety and the kind of symptoms of anxiety may change according to students, students' academic performance suffers from test anxiety (Oetting, 1966).

Test anxiety has a complex structure because it includes both the concept of psychological anxiety and the concept of educational testing. Therefore, a review of the relevant literature reveals the multidimensional structure of test anxiety scales. The Test Anxiety Questionnaire (TAQ), the first example of the conceptualization of test anxiety, was developed by Mandler and Sarason (1952). TAQ were categorized as people who are low or high anxious. Another conceptualization of test anxiety by Liebert and Morris (1967) includes two dimensions, namely worry and emotionality. Test anxiety has changed and gained new dimensions with research conducted on the axis of emotional and cognitive concepts. According to contemporary research, the extent of test anxiety has also changed with the development of new scales. Zeidner (1998) defined the construct of test anxiety as “the set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or failure on an exam or similar evaluative situation” (p.17). This definition highlights the comprehensive nature of test anxiety, encompassing various aspects such as emotional, physiological, and behavioral responses that individuals experience in assessment situations. In today's research, the scope of test anxiety has changed with this perspective. Simultaneously, test anxiety measurement tools have been developed under the biopsychosocial model, too (Friedman & Bendas-Jacob, 1997; Lowe & Lee, 2008; Wren & Benson, 2004). Accordingly, the model posits that test anxiety is the result of the interaction of biological, psychological, and social influences that shape an individual's experience of anxiety in test situations. In the context of test anxiety, the biopsychosocial model provides a framework for understanding the complex interplay of biological, psychological, and social factors that contribute to an individual's experience of anxiety during test situations (Lowe & Lee, 2008). To illustrate, biological factors may encompass genetic predispositions to anxiety, psychological factors may include cognitive distortions and emotional responses to stress, and social factors may include the influence of parental expectations, peer pressure, and educational environments (von der Embse et al., 2014). The recognition of the interaction of these factors allows for the implementation of more effective strategies to assist in the management of test anxiety (von der Embse et al., 2014). In Türkiye too, it is accepted that test anxiety is a multidimensional structure consisting of cognitive, emotional, and behavioral responses (Kırkık & Balkıs, 2022). In this context, the cognitive dimension of the test anxiety structure represents such aspects as cognitive processes, metacognition, and cognitive restructuring (Matthews et al., 1999; Wong, 2008). In addition, individuals' dysfunctional attitudes, automatic thoughts, and irrational beliefs are considered cognitive dimensions in the test anxiety structure (Wong, 2008). The behavioral dimension of test anxiety covers a series of observable behaviors and reactions that individuals exhibit in evaluation situations as test anxiety has been associated with behavioral symptoms such as avoidance, procrastination, and task-irrelevant behavior (Hembree, 1988). On the other hand, the emotional aspect of test anxiety includes various emotions and emotional reactions that individuals experience in evaluation situations. Test anxiety has been associated with a variety of emotions, including anxiety and emotionality (Wine, 1971).

Test anxiety is therefore linked to the physiological and psychological components experienced during tests and reflects behavioral responses to the emotional and physiological aspects of anxiety.

Test anxiety affects not only individuals' cognitions and emotional reactions but also their academic skills. As various studies have shown, students with high test anxiety have less effective study habits than students with low anxiety (Culler & Holahan, 1980). Poor study skills are also a factor affecting test performance. Other interventions that do not solve the problem of test anxiety will generally be inadequate for students experiencing test anxiety (Sarason & Sarason, 1990). This is consistent with findings suggesting that addressing test anxiety can positively impact academic performance (Brady et al., 2018).

It is important to understand the structure of test anxiety and to review the test anxiety problem in order to resolve test anxiety. Since trying to solve the problem by attributing it to a single cause is not an effective method, it is critical to address the problem from all aspects. Therefore, this present study aims to apply thematic content analysis to test anxiety scales used in Turkish literature and is a scientific resource where researchers working in the field of test anxiety can see the scales that have been developed or adapted on this subject in the last decade. Considering the research results, the compliance of the test anxiety scales with the scale development/adaptation standards was examined. Detailed information is provided on how the structure of test anxiety is addressed and the dimensions of test anxiety scales. The scales developed/adapted in the field of test anxiety in Türkiye in the last decade were examined, considering their weaknesses and strengths. Therefore, it is expected to serve as a scientific resource in the field of test anxiety because it offers up-to-date measurement tools. In this way, it offers researchers working in test anxiety a resource where they can view current scales together.

This study aims to examine the descriptive and psychometric properties of test anxiety scales applied in Türkiye with thematic content analysis. The construct of test anxiety is of great importance for individuals in terms of such factors as academic achievement (Terzi et al., 2019), psychological health (Güngör & Saltürk, 2022), management of cognitive and emotional processes (Çakmak et al., 2017; Gökkaya, 2019), and social support (Kapıkıran, 2020). Considering the relationship between these structures and test anxiety, it would be useful to conduct a thematic content analysis to understand test anxiety. To enable the development of effective strategies to reduce test anxiety and to increase the overall quality of life of individuals, thematic content analysis is necessary for researchers to understand the structure of test anxiety, thereby allowing them to evaluate and manage the effects of test anxiety more effectively. No study in the literature examines test anxiety scales with thematic content analysis. Therefore, the study is considered to be original in this perspective. The data obtained from the study will provide practitioners with a general framework for test anxiety scales to be developed or adapted in the Turkish sample.

Method

The Research Model

In the study, document analysis, one of the qualitative research methods, was conducted for the studies in which test anxiety scales were developed or adapted for the Turkish sample. In qualitative research, various documents from sources provide a wealth of information (Cresswell, 2012).

Data Collection Process

In the study, the developed/adapted scales were searched in Google Scholar, the National Thesis Data Centre of the Council of Higher Education, and TOAD 2.0 (Turkish Measurement Tools Index). Between 2013 and 2023, the year limitation was applied during the browsing process. Measuring the structure of test anxiety is related to the changes in educational policies and practices over the years. The introduction of the centralized examination system in Türkiye has led to an increase in test anxiety (Doğan, 2020). In particular, the “High School Transition Exam” (LGS), introduced in 2014, has become an important stage in determining students’ futures as the competitive environment created by this exam has been a factor that increases anxiety in students (Özcan & Arslan, 2018). The change in the examination system has been effective in our review of the last decade. The keywords [“exam anxiety” OR “test anxiety” AND “scale development” OR “scale validation”] in English and [“sınav kaygısı” AND “ölçek geliştirme” OR “ölçek uyarlama”] in Turkish were used in Google Scholar. There were 3540 studies with English keywords and 624 studies with Turkish keywords. In the National Thesis Data Centre of the Council of Higher Education and TOAD 2.0 index, 207 and 11 studies were examined with the keyword “exam anxiety”, respectively. Thus, a total of 4382 studies were examined. The full texts of all these studies were examined and the studies that developed or adapted the test anxiety scale were included in this particular study. Primary studies that developed/adapted the test anxiety scale for a specific domain (e.g., math test anxiety) were not included in the study. Based on the analysis of the studies, 12 scales were developed or adapted for the test anxiety subject. In one of the studies, it was determined that there was an article related to the thesis, and the relevant study was included in the coding only as an article. In a separate study, it was found that a scale that was developed in Turkish was adapted to the English language, and only the Turkish development study of the scale was included. In addition, another study reviewed was the first adaptation study published by the same author in 2009, and some of the information in the form was taken from this article. As a result, the research was conducted on 10 studies (These studies are presented with an asterisk in the references section). The descriptive statistics of the studies are shown in Table 1.

Table 1

Descriptive Statistics of the Included Studies

	Variables	Frequency
Type of Study	Article	8
	Proceeding	1
	Thesis	1
Year of Study	2013-2015	1
	2016-2018	5
	2019-2021	2
	2022-2024	2
Adaptation or Development	Adaptation	5
	Development	5

Data Collection Tools

Within the scope of the study, the “Scales Used for Test Anxiety Review Form”, which includes the descriptive and psychometric properties of the scales, was created. In the creation of the form, the variables identified in the literature review were taken into account as well as the psychometric properties that the measurement tools should include. The created measurement tool was finalized

after receiving the opinions of three measurement and assessment experts. Three experts were interviewed together and gave feedback on the items whether they were appropriate, inappropriate, or needed to be corrected. In this interview, the experts decided that some items were inappropriate and should be removed and suggested new items. They also reported that some items were appropriate and did not make any corrections. Corrections were made in accordance with the common decision for the items that needed to be corrected. Since the experts made decisions through discussion in this interview, full harmony was achieved. The interview lasted approximately 45 minutes. The items in the review form are listed as follows:

1. Who is the target group of the scales?
2. In the scales
 - a. Has an item pool been created for scale development?
 - b. For adapted scales, which translation method was preferred and what is the distribution of the culture in which it was developed and the source language?
3. Was a pilot study conducted?
4. What are the types of scales used and how do they vary?
5. What is the distribution of the scale number of items?
6. What is the distribution and name of the factors of the scales and the ratio of the number of items to the factors?
7. What are the preferred reliability estimation methods in scales?
8. What are the preferred validity estimation methods in the scales?
9. Were cut-off scores given in the scales?

Data Analysis

Thematic content analysis was conducted in the study. Thematic content analysis, one of the preferred analyses in the literature, is a method frequently favored in articles (Abdul Razak et al., 2023; Gezer, 2022, Şengül-Avşar & Barış-Pekmezci, 2022). In thematic analysis, the researcher identifies themes and discusses the main ideas and results obtained from the studies (Cresswell, 2012). In this study, the initial step was to identify the themes based on various sources, followed by the application of coding to these themes. In the study, coding was carried out by two different researchers. The inter-rater consistency, calculated based on the methodology of Miles and Huberman, was found to be 94.44%. It can be posited that the inter-rater reliability is high. The codings that caused inconsistencies were re-examined by two coders, who then re-coded them to resolve any inconsistencies.

Findings

The findings of the study are provided in this section in line with the items included in the “Scales Used for Test Anxiety Review Form.”

Target Group

The distribution of the target group for which the scales were developed or adapted is given in Table 2.

Table 2

Distribution of the Target Group of Scales

	Categories	Frequency
Target Group	Parents	1
	Primary and Secondary School Students	1
	Secondary School Students	1
	Secondary and High School Students	3
	High School Students	2
	Undergraduate	2

When the target group of the scales in Table 2 is reviewed, it becomes evident that most of the studies were conducted with secondary and high school students. When the scales are analyzed, it is seen that in one study, in contrast to other studies, parents were studied (Baytemir & İlhan, 2018). In addition, a study conducted with secondary school students was stated to be suitable for the target group of 4th grade and above, and data were collected from students preparing for the 8th-grade high school admission exam (Başol, 2017). It was noted that a study conducted with high school students was conducted with senior students preparing for the university entrance examination (Bozkurt et al., 2017).

Creating an Item Pool in Developed Scales

When the scale development studies ($f = 5$) were analyzed, it was observed that an item pool was created in all studies. For example, a 71-item item pool was created for the 18-item “Exam Anxiety Scale for Parents” (Baytemir & İlhan, 2018), while a 29-item item pool was created for the development of the 22-item “State Test Anxiety Scale” (Şahin, 2019).

Translation Method Used in Adaptation Scales, Source Culture, and Source Language

Table 3 presents the distribution of translation methods, source cultures, and source languages used in the studies in which the scales were adapted into Turkish.

Table 3

Translation Method Used in Adapted Scales, Source Culture and Source Language Distribution

	Categories	Frequency
Translation Method	Back Translation	4
	A Single Translation	1
Source Culture	American	1
	Unreported	4
Source Language	English	5

As regards the translation method used in the adapted scales, it was observed that the back translation procedure was used in most of the studies ($f = 4$, 80.0%). Only one study used a single translation procedure and, in this study, the validity and reliability study of the “Westside Test Anxiety Scale” for middle and high school students was conducted (Totan, 2018). The researcher

reported that in the first adaptation study conducted by Totan and Yavuz (2009), no back-translation procedure was performed because the translation was done by experts in the field. When the source cultures in which the adapted scales were developed were examined, it was observed that they were not specified in the majority of the studies ($f = 4, 80.0\%$). Only one scale adaptation study provides information on the source culture. It was noted that the “Cognitive Test Anxiety Scale-Revised” adopted in this study was reported to have been developed in the United States of America (Bozkurt et al., 2017). Considering the language in which the scales were developed in the scale adaptation studies were examined, it was observed that the source language of all scales was English.

Pilot Study

The distribution of pilot studies in scales developed or adapted for test anxiety is presented in Table 4.

Table 4

Distribution of the Status of Pilot Studies

	Categories	Frequency
Pilot Study	Made	7
	Unreported	3

Table 4 shows that pilot studies were conducted for most of the scales ($f = 7, 70.0\%$). There are some studies in which pilot studies are included in the directly browsed studies and one study in which a pilot study was included in another study. It can be said that a pilot study was conducted within the scope of the “Westside Test Anxiety Scale” validity and reliability study conducted by Totan (2018). In the adaptation study of this scale (Totan & Yavuz, 2009), it was indicated that a pilot study was conducted with a sample of 20 students.

Types of Scale

A closer look at the scale type of the scales used in the study revealed that all but one of the scales ($f = 9; 90.0\%$) used a Likert-type scale. In one scale, the type of scale used was not mentioned. The distribution of the number of points in these scales in which Likert-type scale was used is given in Table 5.

Table 5

Distribution of Likert Type Scales

	Categories	Frequency
Likert Type	3-point Likert scale	1
	4-point Likert scale	3
	5-point Likert scale	5

Table 5 shows that most of the scales were developed on a 5-point Likert scale. The next Likert type is the 4-point Likert scale. None of the scales has more than five scale points. Only one scale was developed as a 3-point Likert scale (Tan, 2020) and this scale is designed for middle and high school students.

Number of Items

The distribution of the number of items for the scales included in the study is given in Table 6.

Table 6

Distribution of the Number of Items of the Scales

	Categories	Frequency
Number of Items	11-15 items	2
	16-20 items	3
	21-25 items	3
	26-30 items	2

Table 6 shows that when the item distributions of the scales are analyzed, it is observed that there is no scale with more than 30 items. The two scales with the highest number of items have 30 items each (Aydın & Bulgan, 2017; Tan, 2020). The scale with the minimum number of items is the “Westside Test Anxiety Scale”, the validity study of which was conducted by Totan (2018), and this scale consists of 11 items. This scale is followed by the “IDA Test Anxiety Scale” developed by Başol (2017) with 15 items.

Distribution and Labeling of Factors and the Ratio of the Number of Items to the Number of Factors

Table 7 presents the distribution and labels of the factors, the number of items, and the dimension ratio in the scales considered within the scope of the research.

Table 7

Distribution and Labeling of Dimensions with Number of Factors

Name of Scale	Number of Factors	Naming of dimensions (number of items in the relevant factor)
Exam Anxiety Scale for Parents	2	Worry (9), Psychological (9)
Test Anxiety Scale for Online Exams	2	Physiological and Psychological Anxiety (12), Technical Anxiety (8)
IDA Test Anxiety Scale	2	Cognitive and Physiological Anxiety (9), Family and Surrounding Anxiety (6)
State Test Anxiety Scale	3	Cognitive (9), Psychosocial (5), Physiological (8)
Cognitive Test Anxiety Scale– Revised	1	Cognitive (23)
Test Anxiety Scale for Children and Adolescents	3 and 4	For the three sub-dimensional model; Biological (10), Psychological (10), Social (10) For the four sub-dimensional model; Cognitive (3), Physiological (10), Emotional, (7) Behavioral (10)
Children’s Test Anxiety Scale	3	Thoughts (13), Off-Task Behaviors (8), Autonomic Reactions (9)
Friedben Test Anxiety Scale	3	Cognitive Obstruction, Tenseness, Social Derogation (The number of items is not mentioned)
Multidimensional Exam Anxiety Scale	4	Anxiety (4), Cognitive intervention (4), Tension (4), Physiological indicators (4)
The Westside Test Exam Scale	1	Test anxiety (11)

Table 7 shows that the number of dimensions of the scales varies between 1 and 4. The dimension names of the scales differ from each other. It is observed that there are at least 3 items in the dimensions and 23 items in the scale determined as a one-dimension scale. In addition, the number of items of sub-dimensions was not represented in one scale. It was also determined that some of the dimensions in the scales were named in common. Another remarkable result in the findings is

that in the “Test Anxiety Scale for Children and Adolescents”, both 3-dimensional and 4-dimensional structures were examined on the same data, and different labels were used for both structures.

Reliability Estimation Method

The distribution of the coefficients calculated to provide evidence for the reliability of the scales within the scope of the research is presented in Table 8.

Table 8

Distribution of Reliability Coefficient Types

	Categories	Frequency
Reliability Coefficient	Cronbach's alpha	10
	Composite reliability	1
	Spearman-Brown Split half reliability	2
	Test-retest	4

Table 8 shows that Cronbach's alpha internal consistency reliability coefficient was calculated the most. Secondly, test-retest (TRT) reliability, which is used to determine the form's consistency over time, was used. The composite reliability coefficient was calculated together with Cronbach's alpha in one study (Dikmen, 2023).

Table 9 demonstrates the reliability values obtained for the overall and sub-dimensions of the scales. Table 9 shows that Cronbach's alpha coefficients of the scales vary between 0.86 and 0.94. When the highest reliability value is examined, it becomes evident that it is obtained from the “Test Anxiety Scale for Online Exams” and its subscales. It can be stated that the internal consistency value is high in all scales. It is seen that the TRT reliability value varies between 0.72 and 0.81. The TRT reliability value was calculated together with Cronbach's alpha in the scales and was generally lower than Cronbach's alpha. From this point of view, it was determined that the consistency between items was higher than the consistency against time in the overall scale and sub-dimensions where both reliability coefficients were calculated. The original version of the Cognitive Test Anxiety Scale, developed by Cassady and Johnson (2002), reported a Cronbach's alpha of 0.91, indicating high reliability for the scale. Wren and Benson (2004), who developed Children's Test Anxiety Scale (CTAS), reported a Cronbach's alpha of 0.92 for the scale, demonstrating excellent internal consistency. Friedben Test Anxiety Scale (FTAS) has shown Cronbach's alpha ranging from 0.85 to 0.90 in various studies, indicating good reliability (von der Embse et al., 2014). The overall internal reliability coefficient of the Multidimensional Test Anxiety Scale was obtained as 0.88 (Putwain et al., 2020). The alpha value was not reported in the study in which the Westside Test Anxiety Scale was developed (Driscoll, 2007).

Table 9

General and Sub-dimensional Reliability Coefficient Values for The Scales

Name of Scale	General Reliability Coefficient	Sub-dimensional Reliability Coefficient
Exam Anxiety Scale for Parents	$\alpha = 0.91$ for the first study, $\alpha = 0.93$ for the second study	Worry ($\alpha = 0.88$) and Physiological dimension ($\alpha = 0.91$) for both first and second study
Test Anxiety Scale for Online Exams	$\alpha = 0.94$	Physiological and Psychological anxiety ($\alpha = 0.93$) Technical anxiety ($\alpha = 0.90$); Physiological and Psychological anxiety (CR = 0.94), Technical anxiety (CR = 0.90)
IDA Test Anxiety Scale	$\alpha = 0.86$ $\rho = 0.797$; $r_{TRT} = 0.811$	Cognitive and Physiological Anxiety ($\alpha = 0.89$), Family and Surrounding Anxiety ($\alpha = 0.70$)
State Test Anxiety Scale (STAS)	$\alpha = 0.94$; $\rho = 0.96$; $r_{TRT} = 0.81$	Cognitive ($\alpha = 0.93$), Psychosocial ($\alpha = 0.84$), Physiological ($\alpha = 0.85$); Cognitive ($\rho = 0.89$), Psychosocial ($\rho = 0.86$), Physiological ($\rho = 0.84$); Cognitive ($r_{TRT} = 0.74$), Psychosocial ($r_{TRT} = 0.80$), Physiological ($r_{TRT} = 0.78$)
Cognitive Test Anxiety Scale– Revised	$\alpha = 0.93$; $r_{TRT} = 0.93$	-
Test Anxiety Scale for Children and Adolescents	$\alpha = 0.91$	Biological ($\alpha = 0.83$), Psychological ($\alpha = 0.85$), Social ($\alpha = 0.83$)
Children’s Test Anxiety Scale	$\alpha = 0.88$	Thoughts ($\alpha = 0.82$), Off-Task Behavior ($\alpha = 0.72$), Autonomic Reactions ($\alpha = 0.75$)
Friedben Test Anxiety Scale	$\alpha = 0.88$	Cognitive Obstruction ($\alpha = 0.76$), Tenseness ($\alpha = 0.81$), Social Derogation ($\alpha = 0.78$)
Multidimensional Exam Anxiety Scale	$\alpha = 0.86$	Anxiety ($\alpha = 0.84$), Cognitive Intervention ($\alpha = 0.82$), Tension ($\alpha = 0.83$), Physiological indicators ($\alpha = 0.78$)
The Westside Test Exam Scale	For middle school students $\alpha = 0.92$, For high school students $\alpha = 0.91$; For middle school students $r_{TRT} = 0.72$, For high school students $r_{TRT} = 0.74$	-

Note. α = Cronbach’s alpha reliability coefficient, ρ = Spearman-Brown split half reliability coefficient, r_{TRT} = Test-retest reliability coefficient, CR = Composite reliability coefficient.

Validity Evidence

The distribution of the coefficients calculated to provide evidence for the validity of the scales within the scope of the research is presented in Table 10.

Table 10

Distribution of Validity Evidence

Validity	Categories	Frequency
Construct Validity	Exploratory Factor Analysis	7
	Confirmatory Factor Analysis	9
Criterion-Related Validity	Convergent Validity	6
	Discriminant Validity	1
Content Validity		0

As can be observed in Table 10, construct validity was examined most frequently in the scales, while criterion validity was the second most frequently examined. Content validity was not examined in the studies. The findings regarding EFA and CFA results in the studies in which construct validity was examined are presented in Table 11. In addition, Table 12 presents the scales used in the studies in which criterion validity was examined.

Table 11

EFA (explained variance) and CFA (fit indices) Results of The Measurement Tools

Name of Scale	Explained Variance	Criteria
Exam Anxiety Scale for Parents	55.44%	$\chi^2/df=2.02$, RMSEA=.058, SRMR=.075, CFI=.98, NFI=.97, TLI=.98
Test Anxiety Scale for Online Exams	63.4%	$\chi^2/df=3.242$, RMSEA=.075, SRMR=.062, CFI=.89, GFI=.87, TLI=.87
IDA Test Anxiety Scale	53,73%	$\chi^2/df = 3.06$, RMSEA=.070, NNFI=.96, CFI=.97, GFI=.92, AGFI=.89
State Test Anxiety Scale	59.21%	$\chi^2/df=1.72$, CFI=.96, NNFI=.96, IFI=.96, RMSEA=.05, SRMR=.05
Cognitive Test Anxiety Scale-Revised	38.8%	CFI=.984, TLI=.982, RMSEA=.044, SRMR=.056
Test Anxiety Scale for Children and Adolescents	Three dimensions: 43.00% Four dimensions: model: 47.64%	-
Children's Test Anxiety Scale	-	$\chi^2=1414.72$, $df=356.35$, $p=.00$, $\chi^2/df = 3.97$, RMSEA=.05, SRMR=.05, CFI=.97, GFI=.92, AGFI=.90
Friedben Test Anxiety Scale	-	$\chi^2=237.603$, $df=95$, RMSEA=.05, CFI=.94, TLI=.92
Multidimensional Exam Anxiety Scale	-	$\chi^2= 669.27$, $df=216$, RMSEA=.071, CFI=.87, GFI=.88, IFI=.88, SRMR=.073
The Westside Test Exam Scale	55.00%	Secondary school: $\chi^2= 137.96$, $df=44$, $\chi^2/df= 3.14$, CFI=.97, SRMR=.003 High school: $\chi^2=128.33$, $df=44$, $\chi^2/df=2.92$, CFI=.96, SRMR=.003

When Table 11 is examined, both EFA and CFA analyses were conducted for six scales. Only EFA was reported in one study and CFA was reported in three studies. When the explained variance values are examined, it is seen that the lowest explained variance was determined in the Cognitive Test Anxiety Scale-Revised form (Bozkurt et al., 2017). However, since the scale is unidimensional, the explained variance ratio obtained is sufficient. It can be stated that the fit indices obtained according to the CFA results in all scales generally show acceptable and excellent fit. As a result, it can be interpreted that the scales considered within the scope of the research represent the relevant

structure. While the construct validity was examined with CFA for the Multidimensional Exam Anxiety Scale (Ergin & Ekşi, 2022), the Friedben Test Anxiety Scale (Akın et al., 2013) and the Children's Test Anxiety Scale (Aydın & Bulgan, 2017), EFA and CFA were used to examine the construct validity in other scales.

Table 12

Distribution of Criterion Validity Evidence

Name of Scale	Criteria
Exam Anxiety Scale for Parents	"State and Trait Anxiety Scales"
Test Anxiety Scale for Online Exams	Convergent validity: Factor loadings and AVE, Discriminant validity: Fornell-Larcker criterion and Heterotrait-Monotrait
IDA Test Anxiety Scale	"Spielberger Test Anxiety Scale" and "Westside Test Anxiety Scale"
State Test Anxiety Scale	"The Revised Test Anxiety Scale"
Cognitive Test Anxiety Scale-Revised	"Test Anxiety Inventory (TAI)"
Test Anxiety Scale for Children and Adolescents	-
Children's Test Anxiety Scale	-
Friedben Test Anxiety Scale	-
Multidimensional Exam Anxiety Scale	"The Revised Test Anxiety Scale"
The Westside Test Exam Scale	-

As illustrated in Table 12, the criterion validity was examined in six studies. In all studies, both construct validity and criterion validity were examined together. For only one scale, different criteria were used for both convergent and discriminant validity (Dikmen, 2023). For the other scales, only convergent validity analyses were conducted.

Cut off Score

Only one scale has a cut-off score. If the total score obtained from the IDA Test Anxiety Scale is below 45 points, there is no test anxiety at a level of concern. Scores between 45-60 indicate moderate test anxiety and those above 60 indicate high test anxiety (Başol, 2017). There is no cut-off score in the other nine scales.

Discussion

In this study, the scales developed or adapted to measure the construct of test anxiety in the Turkish sample were examined in the context of various variables. The first aim of the study was to examine the characteristics of the group to which the scales were applied. The target groups for test anxiety scales vary widely, encompassing different age ranges, educational levels, and contexts. Commonly, the scales target students, ranging from elementary to university levels, as they are frequently exposed to academic evaluations and can experience test anxiety (Cassady & Johnson, 2002; Zeidner, 1998). In Türkiye, test anxiety intensifies during certain periods due to the structure of the education system and its exam-focused approach. Especially since high school entrance exams are applied after secondary school and university entrance exams are in question after high school, the examination of exam anxiety of these student groups has been intensive. When the literature is examined, it is seen that the age periods when test anxiety is most effective are 11-14 years old. It is stated that test anxiety, which is reported as 40% in the young population, varies between 20% and 35% in all university students (McDonald, 2001; Trigueros et al., 2020). Similarly, the structure of test anxiety

was also studied in the data obtained from university students, as there are exams for recruitment after university in Türkiye.

Almost all the scales are of Likert type, with the 5-point Likert scale being the most preferred one. The 5-point Likert scale has its simplicity and ease of use. This simplicity enhances the response rate and the quality of the data collected. In the context of Türkiye, the 5-point Likert scale has been found to be culturally appropriate and effective. Turkish students are accustomed to Likert-type scales in educational assessments, making it a familiar and accepted tool for measuring psychological constructs such as test anxiety (Tezbaşaran, 1997).

Although the number of items on the scale is at least 11 items and at most 30 items, it is mostly in the range of 16-25 items. Items variation allows for both quick assessments and in-depth evaluations depending on the research or clinical need (Sarason, 1984; Spielberger et al., 1980). When the number of dimensions of the scales is analyzed, it has been observed that a 3-dimensional structure is mostly adopted. Although the names of the factor structures of the scales differ in each of them, the dimensions named psychological, physiological, cognitive, tension, and anxiety were addressed in more than one scale with common names. It can be observed that test anxiety scales contain multidimensional structures and generally include cognitive and physiological factors (Friedman & Bendas-Jacob, 1997; Lowe & Lee, 2008; Sarason & Sarason, 1990; Wren & Benson, 2004). The literature shows that the symptoms of test anxiety are classified under four main dimensions: cognitive, physiological, behavioral, and emotional symptoms (Yeşilyurt, 2007). Therefore, when determining the factor structure of test anxiety scales, all relevant dimensions should be taken into consideration. Looking at the number of items in the dimensions, three of the scales have a consistent number of items in the sub-dimensions, while the number of items varies in the sub-dimensions of the other scales. The proportion of items varies across factors; some scales have a balanced distribution, while others emphasize certain dimensions more. For instance, the Cognitive Test Anxiety Scale (CTAS) emphasizes cognitive aspects of test anxiety, as evidenced by both its factor structure and item distribution (Cassady & Johnson, 2002), like the Turkish version (Bozkurt et al., 2017). The distribution of the number of items in the scales varies, with some scales containing a higher number of items to capture different facets of test anxiety comprehensively (Cassady & Johnson, 2002).

In all the scale development studies examined, item pools were created to ensure that the items comprehensively and accurately cover the construct of test anxiety and are relevant to the target group. The development and analysis of an item pool are critical steps in creating effective test anxiety scales. Researchers should consider the comprehensive content, expert opinions and culture when developing item pools. Thus, by developing test anxiety scales with high validity, they will provide accurate information for educational practices and policies. Therefore, it is of the utmost importance to conduct pilot studies on the item pool to ascertain its performance. The lack of pilot studies poses significant potential risks to the validity and reliability of test anxiety scales. While the majority of the reviewed studies conducted pilot studies to test the initial versions of the scales, improved the scales based on feedback, and examined the appropriateness of the scales for measuring test anxiety, others lack the required information on whether pilot studies were conducted. Pilot testing often involves small, representative samples and provides crucial feedback for scale modifications (Cassady & Johnson, 2002). It is, therefore, important to note that the risk of reaching erroneous results increases in scales that are not piloted because the preliminary stages of validity testing will have been ignored. Likewise, reliability is also affected by this situation and pilot studies often provide data on reliability. For the abovementioned reasons, the lack of pilot

studies should be considered a major methodological deficiency in a scale and should be seen as a fundamental factor affecting the reliability and validity of measures.

In scale adaptation, particularly in studies involving diverse populations, it is important to ensure cultural relevance and linguistic equivalence at the translation stage (Furlan et al., 2009). For this purpose, translation methods such as single translation, back translation and committee approach methods are commonly used (Brislin, 1970). The back translation method was used in all but one of the scales. Regarding the distribution of the cultures in which the scales were developed, information is given for one scale, which was adapted from American culture (Bozkurt et al., 2017), and the language of all scales developed as the source language was English.

Cronbach's alpha coefficient, one of the reliability estimation methods, was reported in all of the test anxiety scales. In addition, test-retest reliability coefficient and Spearman-Brown split-half reliability method were also reported in some scales. Cronbach's alpha calculated for all scales examined shows high internal consistency (typically above 0.70) (Nunnally & Bernstein, 1994). Test constructors for test anxiety scales have generally produced measures with high-reliability scores of 0.80 and 0.90 (Zeidner, 1998). When the reliability values of the original version of the scales are analyzed, there are generally no great differences in the reliability scores of the adapted scales. This helps us to conclude that there should be no concern about the reliability scores of the adapted scales. Reliability estimation methods ensured that the scales produced consistent and stable results over time and across different populations (Nunnally & Bernstein, 1994). Although composite reliability, like Cronbach's alpha, assesses the reliability of a multi-item scale by considering the factor loadings and error variances of the items, it was calculated for only one of them in the scales reviewed. In Turkish test anxiety research, composite reliability is increasingly reported, especially in studies employing confirmatory factor analysis (CFA) to validate the scale's structure.

When the scales subject to the research were analyzed in terms of validity evidence, construct validity was the most considered validity for the scales. Construct validity is about how well a measurement tool measures the targeted construct (Haynes et al., 1995). In the reviewed studies, CFA was generally conducted together with exploratory factor analysis (EFA). While EFA provides initial insights into the structure of the scale, and CFA confirms the factor structure and relationships between variables, ultimately ensuring the validity of the measurement instrument (Marsh et al., 2014). EFA typically requires more data and complex analysis, while CFA provides faster results with less data. In scale adaptation studies, since the scale was previously developed in another culture and the factor structure was determined, it is generally seen that only CFA is applied without EFA. While some researchers find the exploratory information provided by EFA unnecessary and adopt a direct confirmatory approach (Orçan, 2018), others emphasize that using EFA and CFA together yields better results (Marsh et al., 2014). It may be reasonable for the researchers to examine the scale only with CFA to assess the appropriateness of culture in which it was developed for saving time and resources. Content validity determines how well a measurement tool represents the concept it is intended to measure. Collecting evidence of content validity shows that the construct is addressed in a comprehensive approach. The lack of stated content validity on whether in the examined studies shows that sufficient evidence regarding validity is not provided. While most researchers focus on tests of validity and reliability, they ignore more subjective measures such as content validity in reporting. Since the researchers did not include it in their studies, it cannot be predicted to what extent content validity taken into account. If a scale without content validity is used, invalid measurements can result in incorrect conclusions and misleading data, which can, in

turn, lead to misdirected educational policies and intervention programs. In addition, the lack of sufficient understanding of the psychological and social factors that affect students' test anxiety can have negative consequences in terms of managing this situation. Therefore, it is expected to provide knowledge of content validity for the scales developed.

When scale development and adaptation studies are analyzed in terms of criterion validity, convergent validity is assessed by examining the correlation between test anxiety scores and measures of similar constructs such as general anxiety or performance anxiety. High correlations provide evidence that the test anxiety scale measures a construct theoretically related to other forms of anxiety. On the other hand, discriminant validity involves demonstrating that the test anxiety scale is not correlated with measures of other constructs such as depression or self-esteem. Low correlations with these unrelated constructs provide evidence that the test anxiety scale is distinct and measures a unique construct (Campbell & Fiske, 1959). Rosas and Furlan (2017) emphasize the importance of establishing the convergent and divergent criterion validity of test anxiety scales to ensure that they accurately measure the intended construct. The validity process involves confirming the validity of the scale in assessing test anxiety by demonstrating that it correlates as expected with relevant constructs and differs from irrelevant constructs.

In scale adaptation and development studies, only one scale was found to have a cut-off score. When literature is examined, the presence of cut-off points in test anxiety scales varies across studies. Scales that provide cut-off points often do so to facilitate the identification of individuals with clinically significant levels of test anxiety. For instance, the Test Anxiety Inventory (TAI) includes cut-off points that categorize individuals into low, moderate, and high anxiety levels based on their scores (Spielberger, 1980). Cut points on test anxiety scales evaluate test anxiety levels effectively while considering the trade-off between sensitivity and specificity, thereby ensuring accurate identification without overlooking or misclassifying cases. On the other hand, some scales do not offer cut-off points, focusing instead on providing a continuous measure of test anxiety. These scales argue that test anxiety is better understood as a spectrum rather than a discrete category (Zeidner, 1998). This approach facilitates understanding individuals' anxiety profiles and creating personalized intervention plans according to their needs. Cut-off scores may not always be valid in different sample groups due to cultural differences. Therefore, assessments without a cut-off score may provide more compatible and flexible results in different demographic groups. When continuous data are transformed into categorical data in measurements with cut-off points, it becomes difficult to interpret the data in detail in the longitudinally. On the other hand, in measurements without a cut-off score, it is possible to better monitor changes over time. For example, an increase or decrease in an individual's anxiety level can be more easily observed in the score distribution, which provides an advantage in evaluating the dynamic change process. The presence of cut-off points of scales facilitates clinical or practical applications determining the severity of anxiety and guiding appropriate interventions based on the identified anxiety levels while others maintain a continuous measurement approach for a more comprehensive assessment.

Data should be collected from participants of different ages and education levels in order to determine the effect of test anxiety on various age groups. The same scale may need to be used when conducting comparative studies to explain the variables affecting test anxiety in students of different age groups. To ascertain the suitability of the scales for use with different groups, it would be beneficial to conduct further analyses with participants who possess varying socio-economic and demographic characteristics. For example, validity studies can be conducted for scales across a wide age range, extending beyond high school and university levels to encompass the primary school

level as well. The determination of test anxiety at an early age can provide crucial data on the evolution of this anxiety over time, thus facilitating the implementation of preventive measures in educational processes. In addition, researchers may develop more comprehensive scales by considering factors such as demographic characteristics, academic background, and social support. Considering the multidimensional nature of test anxiety in these scales proposed to be developed will enable the structure to be assessed comprehensively. Due to the nature of this study in which thematic content analysis was used, it may contain subjectivity since the evaluation criteria were determined by the researchers. In future research, the psychometric properties of other scales in the literature can be examined using the evaluation criteria in this study or other criteria in addition to these criteria.

Ethics Committee Approval: Since this study does not fall within the scope of human/animal research and the data consists of primary studies, it is not within the scope of the ethics committee.

Author Contributions: The authors contributed equally to the study.

Conflict of Interest: There is no conflict of interest between the authors.

Sınav Kaygısı Ölçeklerinin Tematik İçerik Analizi ile İncelenmesi

Nurevsa Münevver Bahçeci^a  Yıldız Yıldırım^b  Melek Gülşah Şahin^c 

^a Psikolog, Artı Enstitü, Ankara, Türkiye, nmbahceci@gmail.com

^b Dr. Öğr. Üyesi, Aydın Adnan Menderes Üniversitesi, Aydın, Türkiye, yildizyldrm@gmail.com

^c Prof. Dr., Gazi Üniversitesi, Ankara, Türkiye, mgulsahsahin@gazi.edu.tr

ÖZET

Bu çalışmanın amacı sınav kaygısı için kullanılan ölçekleri incelemektir. Türkiye'de geliştirilen veya uyarlanan sınav kaygısı ölçekleri tematik içerik analizi ile incelenmiştir. 2013 yılından bu yana yapılan çalışmalar incelenerek 12 ölçek elde edilmiş, 2 çalışma bazı kriterleri karşılamadığı için değerlendirmeye alınmamıştır. Veri toplama yöntemi olarak doküman analizi kullanılmıştır. İncelenen ölçeklerin büyük çoğunluğu makalelerden oluşmaktadır. Ölçek geliştirme ve uyarlama çalışmaları çoğunlukla 2016-2018 yılları arasında yapılmıştır. Ölçeklerin çoğu Likert tipindedir. Madde sayıları 11 ila 30 arasında değişmektedir. Bazı ölçeklerde sınav kaygısının çok boyutlu yapısı dikkate alınmamıştır. Sınav kaygısının tek boyutlu yapısını ölçen ölçekler bulunmaktadır. Güvenirlik tahminlerinde tüm ölçekler için Cronbach alfa katsayısı hesaplanmış ancak bileşik güvenilirlik, Spearman-Brown iki yarı güvenilirliği ve test-tekrar test gibi farklı güvenilirlik kestirim yöntemleri de kullanılmıştır. Cronbach alfa katsayısı minimum 0,70 ve maksimum 0,94 olarak hesaplanmıştır. Ölçeklerin geçerliğine ilişkin veriler incelendiğinde en sık yapı geçerliğine bakıldığı, bunu ölçüt geçerliğinin izlediği görülmektedir. Çalışmalarda kapsam geçerliği incelenmemiştir. Yapı geçerliliğini inceleyen çalışmalarda AFA ve DFA değerlerine, ölçüt geçerliğini inceleyen çalışmalarda ise kullanılan ölçeklere ilişkin bulgular sunulmuştur. Aynı zamanda ölçeklerin psikometrik özelliklerinin analizinde bazı eksiklikler tespit edilmiş ve bu doğrultuda önerilerde bulunulmuştur.

MAKALE BİLGİSİ

Makale Türü

Araştırma

Makale Geçmişi

Gönderim tarihi:

02.08.2024

Kabul tarihi:

03.11.2024

Anahtar Kelimeler

Sınav Kaygısı, Ölçek Geliştirme, Ölçek Uyarlama, Tematik İçerik Analizi, Psikometrik Özellikler

Atıf Bilgisi: Bahçeci, N. M., Yıldırım, Y., ve Şahin, M. G. (2024). Sınav kaygısı ölçeklerinin tematik içerik analizi ile incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 12 (3), 1513-1548. <https://doi.org/10.46778/goputeb.1526461>

Sorumlu yazar: Yıldız Yıldırım, e-posta: yildizyldrm@gmail.com

Giriş

Sınav kaygısı, ilkokuldan lisansüstü eğitime kadar farklı eğitim seviyelerindeki öğrencileri etkileyen yaygın bir sorundur (Chapell ve diğerleri, 2005). Sınav kaygısı çeken bireyler, yaş fark etmeksizin sınavlardan önce, sınav sırasında veya sınavlardan sonra deneyimledikleri huzursuzluk, endişe ve korku duygularını içerir. Kaygının yoğunluğu ve kaygının belirtileri öğrencilere göre değişebilir de öğrencilerin akademik performansı sınav kaygısından etkilenmektedir (Oetting, 1966).

Sınav kaygısı hem psikolojik kaygı kavramını hem de eğitsel test kavramını içerdiği için karmaşık bir yapıya sahiptir. Bu nedenle literatür incelendiğinde sınav kaygısı ölçeklerinin çok boyutlu yapısı görülmektedir. Sınav Kaygısı Anketi (TAQ), sınav kaygısının kavramsallaştırılmasının ilk örneği olup Mandler ve Sarason (1952) tarafından geliştirilmiştir. TAQ düşük ve yüksek kaygılı kişiler olarak kategorize edilmiştir. Liebert ve Morris (1967) tarafından yapılan bir diğer sınav kaygısı kavramı ise endişe ve duygusallık olmak üzere iki boyutu içermektedir. Duygusal ve bilişsel kavramlar ekseninde yapılan araştırmalarla sınav kaygısı değişmiş ve yeni boyutlar kazanmıştır. Çağdaş araştırmalara göre sınav kaygısının kapsamı da yeni ölçeklerin geliştirilmesiyle değişmiştir. Zeidner (1998, s.17)'e göre sınav kaygısı yapısı "bir sınavda veya benzer değerlendirme durumlarında olası olumsuz sonuçlar veya başarısızlık hakkındaki endişeye eşlik eden fenomenolojik, fizyolojik ve davranışsal tepkiler kümesi" olarak tanımlanmıştır. Bu tanım, sınav kaygısının kapsamlı doğasını vurgular ve bireylerin değerlendirme durumlarında deneyimledikleri duygusal, fizyolojik ve davranışsal tepkiler gibi çeşitli yönleri kapsar. Günümüz araştırmalarında, sınav kaygısının kapsamı bu bakış açısıyla değişmiştir. Aynı zamanda, sınav kaygısı ölçüm araçları da biyopsikososyal modele uygun olarak geliştirilmiştir (Friedman ve Bendas-Jacob, 1997; Lowe ve Lee, 2008; Wren ve Benson, 2004). Buna göre model, sınav kaygısının, bireyin sınav durumlarında yaşadığı kaygı deneyimini şekillendiren biyolojik, psikolojik ve sosyal etkilerin etkileşiminin bir sonucu olduğunu ileri sürmektedir. Sınav kaygısı bağlamında biyopsikososyal model, bireyin sınav durumlarında yaşadığı kaygıya katkıda bulunan biyolojik, psikolojik ve sosyal faktörlerin karmaşık etkileşimini anlamak için bir çerçeve sunmaktadır (Lowe ve Lee, 2008). Örneğin, biyolojik faktörler kaygıya yönelik genetik yatkınlıkları kapsayabilirken, psikolojik faktörler bilişsel çarpıtmaları ve strese karşı duygusal tepkileri içerebilir. Sosyal faktörler ebeveyn beklentilerinin, akran baskısının ve eğitim ortamlarının etkisini içerebilir (von der Embse ve diğerleri, 2014). Bu faktörlerin etkileşiminin tanınması, sınav kaygısının yönetilmesine yardımcı olmak için daha etkili stratejilerin uygulanmasına olanak tanır (von der Embse ve diğerleri, 2014). Türkiye'de de sınav kaygısının bilişsel, duygusal ve davranışsal tepkilerden oluşan çok boyutlu bir yapı olduğu kabul edilmektedir (Kırkık ve Balkıs, 2022). Bu bağlamda, sınav kaygısı yapısının bilişsel boyutu; bilişsel süreçler, üst biliş ve bilişsel yeniden yapılandırma gibi yönleri temsil eder (Matthews ve diğerleri, 1999; Wong, 2008). Ayrıca, bireylerin işlevsiz tutumları, otomatik düşünceleri ve irrasyonel inançları sınav kaygısı yapısında bilişsel boyutlar olarak ele alınır (Wong, 2008). Sınav kaygısının davranışsal boyutu, bireylerin değerlendirme durumlarında sergiledikleri bir dizi gözlemlenebilir davranış ve tepkiyi kapsar. Sınav kaygısı, kaçınma, erteleme ve görevle ilgisiz davranış gibi davranışsal semptomlarla ilişkilendirilmiştir (Hembree, 1988). Sınav kaygısının duygusal yönü, bireylerin değerlendirme durumlarında deneyimledikleri çeşitli duyguları ve duygusal tepkileri içerir. Sınav kaygısı, kaygı ve duygusallık da dâhil olmak üzere çeşitli duygularla ilişkilendirilmiştir (Wine, 1971). Özetle, sınav kaygısı, testler sırasında deneyimlenen fizyolojik ve psikolojik bileşenlerle bağlantılıdır ve kaygının duygusal ve fizyolojik yönlerine verilen davranışsal tepkileri yansıtır.

Sınav kaygısı yalnızca bireylerin bilişlerini ve duygusal tepkilerini değil, aynı zamanda akademik becerilerini de etkiler. Çeşitli araştırmaların gösterdiği gibi, yüksek sınav kaygısı olan öğrencilerin düşük kaygısı olan öğrencilere göre daha az etkili çalışma alışkanlıkları vardır (Culler ve Holahan, 1980). Zayıf çalışma becerileri de sınav performansını etkileyen bir faktördür. Sınav kaygısı sorununu çözmeyen diğer müdahaleler genellikle sınav kaygısı yaşayan öğrenciler için yetersiz kalacaktır (Sarason ve Sarason, 1990). Bu, sınav kaygısını ele almanın akademik performansı olumlu yönde etkileyebileceğini öne süren bulgularla tutarlıdır (Brady ve diğerleri, 2018).

Sınav kaygısını çözebilmek için sınav kaygısının yapısını anlamak ve sınav kaygısı sorununu incelemek önemlidir. Sorunu tek bir nedene bağlayarak çözmeye çalışmak etkili bir yöntem olmadığından sorunu tüm yönleriyle ele almak kritik öneme sahiptir. Bu nedenle bu çalışmada Türkçe literatürde kullanılan sınav kaygısı ölçeklerine tematik içerik analizi uygulanması amaçlanmıştır. Bu çalışma sınav kaygısı alanında çalışan araştırmacıların son on yılda konu hakkında geliştirilen veya uyarlanan ölçekleri görebilecekleri bilimsel bir kaynak niteliğindedir. Araştırma sonuçları dikkate alındığında sınav kaygısı ölçeklerinin ölçek geliştirme veya uyarlama standartlarına uygunluğu incelenmiştir. Sınav kaygısının yapısının nasıl ele alındığı ve sınav kaygısı ölçeklerinin boyutları hakkında ayrıntılı bilgi verilmiştir. Türkiye'de son on yılda sınav kaygısı alanında geliştirilen veya uyarlanan ölçekler zayıf ve güçlü yönleri dikkate alınarak incelenmiştir. Bu nedenle güncel ölçme araçları sunması nedeniyle sınav kaygısı alanında bilimsel bir kaynak görevi görmesi beklenmektedir. Bu şekilde sınav kaygısı alanında çalışan araştırmacılara güncel ölçekleri bir arada görebilecekleri bir kaynak sunmaktadır.

Bu çalışmanın amacı, Türkiye'de uygulanan sınav kaygısı ölçeklerinin betimsel ve psikometrik özelliklerini tematik içerik analizi ile incelemektir. Sınav kaygısı yapısı, akademik başarı (Terzi ve diğerleri, 2019), psikolojik sağlık (Güngör ve Saltürk, 2022), bilişsel ve duygusal süreçlerin yönetimi (Çakmak ve diğerleri, 2017; Gökkaya, 2019) ve sosyal destek (Kapıkıran, 2020) gibi birçok faktör açısından bireyler için büyük önem taşımaktadır. Bu yapıların sınav kaygısı ile ilişkisi göz önüne alındığında sınav kaygısını anlamak için tematik içerik analizi yapılması yararlı olacaktır. Sınav kaygısını azaltmaya yönelik etkili stratejilerin geliştirilmesine olanak sağlamak ve bireylerin genel yaşam kalitelerini artırmak amacıyla araştırmacıların sınav kaygısının yapısını anlayarak sınav kaygısının etkilerini daha iyi değerlendirip yönetebilmeleri için tematik içerik analizi yapılması gerekmektedir. Literatürde sınav kaygısı ölçeklerini tematik içerik analizi ile inceleyen bir çalışmaya rastlanmamıştır. Bu nedenle çalışmanın bu açıdan özgün olduğu düşünülmektedir. Çalışmadan elde edilen veriler uygulayıcılara Türk örnekleminde geliştirilen veya uyarlanan sınav kaygısı ölçekleri için genel bir çerçeve sunacaktır.

Yöntem

Araştırma Modeli

Bu çalışmada Türk örnekleminde uygulanan sınav kaygısı ölçeklerinin geliştirildiği ya da uyarlandığı araştırmalara yönelik nitel araştırma yöntemlerinden biri olan doküman analizi yapılmıştır. Nitel araştırmalarda, çeşitli kaynaklardan gelen dokümanlar zengin bir bilgi kaynağı sağlar (Cresswell, 2012).

Veri Toplama Süreci

Çalışmada 2013-2023 yılları arasında Google Akademik, YÖK Ulusal Tez Merkezi, TOAD 2.0'da (Türkiye Ölçme Araçları Dizini) taranan çalışmalar incelenmiştir. Sınav kaygısının yapısının

ölçülmesi yıllar içinde değişen eğitim politikaları ve uygulamaları ile ilişkilidir. Türkiye’de merkezi sınav sisteminin getirilmesi sınav kaygısının artmasına neden olmuştur (Doğan, 2020). Özellikle 2014 yılında uygulamaya konulan “Liseye Geçiş Sınavı” (LGS) öğrencilerin geleceğini belirleyen önemli bir aşama haline gelmiştir. Bu sınavın yarattığı rekabet ortamı öğrencilerde kaygıyı artıran bir etken olmuştur (Özcan ve Arslan, 2018). Sınav sistemindeki değişim son on yıla ilişkin sınav kaygısı ölçeklerinin incelenmesinde etkili olmuştur. Taranan dizinlerden Google Akademik’te İngilizce olarak [“exam anxiety” OR “test anxiety” AND “scale development” OR “scale validation”] Türkçe olarak [“sınav kaygısı” AND “ölçek geliştirme” OR “ölçek uyarlama”] anahtar kelimeleri kullanılmıştır. İngilizce anahtar kelimeler ile 3540, Türkçe anahtar kelimeler ile 624 çalışmaya rastlanmıştır. Ulusal tez merkezinde ve TOAD 2.0 dizininde ise “sınav kaygısı” anahtar kelimesi ile tarama yapılmış ve sırasıyla 207 ve 11 çalışma incelenmiştir. Böylece toplam 4382 çalışma incelenmiştir. Tüm bu çalışmaların tam metinleri incelenmiş ve sınav kaygısı ölçeğini geliştiren veya uyarlayan çalışmalar bu araştırmaya dâhil edilmiştir. Sınav kaygısı ölçeğini belirli bir alan (örneğin; matematik sınav kaygısı) için geliştiren veya uyarlayan birincil çalışmalar çalışmaya dâhil edilmemiştir. İncelenen çalışmalar sonucunda toplam 12 çalışmada sınav kaygısı ölçeğinin geliştirildiği ya da uyarlandığı belirlenmiştir. Bu çalışmalardan bir tanesinde tez ile ilişkili makale olduğu belirlenmiş olup kodlamalara ilgili çalışma sadece makale olarak dâhil edilmiştir. Bu çalışmalardan başka birinde de Türkçe geliştirilen bir ölçeğin İngilizceye uyarlandığı görülmüş ve sadece ölçeğin Türkçe geliştirilme çalışması dâhil edilmiştir. Ayrıca incelenen çalışmalardan bir diğerinde aynı yazarın 2009 yılında yayımladığı ilk uyarlama çalışması bulunmaktadır ve bilgilerin bir kısmı bu makaleden alınmıştır. Sonuç olarak araştırma 10 çalışma üzerinden yürütülmüştür. (Bu çalışmalar kaynakça bölümünde yıldız işaretiyle gösterilmiştir). Araştırmada ele alınan çalışmalara ilişkin betimsel özellikler Tablo 1 de sunulmuştur.

Tablo 1

Dahil Edilen Çalışmaların Betimsel Özellikleri

	Değişken	Frekans
Çalışma Türü	Makale	8
	Bildiri	1
	Tez	1
Çalışma Yılı	2013-2015	1
	2016-2018	5
	2019-2021	2
	2022-2024	2
Uyarlama veya Geliştirme	Uyarlama	5
	Geliştirme	5

Veri Toplama Araçları

Çalışma kapsamında ölçeklerin betimsel ve psikometrik özelliklerinin yer aldığı “Sınav Kaygısı Ölçeği İnceleme Formu” oluşturulmuştur. Form oluşturulurken literatür taraması ile belirlenen değişkenler dikkate alınmış, ayrıca ölçme araçlarının içermesi gereken psikometrik özellikler göz önünde bulundurulmuştur. Hazırlanan ölçme aracına yönelik üç ölçme ve değerlendirme uzmanından görüş alınarak nihai hali verilmiştir. Üç uzman birlikte görüşme yapmış ve uzmanlar maddelere ilişkin uygun/uygun değil/düzeltilmeli şeklinde geri bildirimde bulunmuşlardır. Bu görüşmede uzmanlar bazı maddelerin uygun olmadığına ve çıkarılmasına karar vermiştir ve yeni madde önerisinde bulunmuşlardır. Bazı maddelerin ise uygun olduğunu belirtmiş ve herhangi bir düzeltme yapmamışlardır. Düzeltmesi gereken maddeler için ise ortak karara göre düzeltmeler

yapılmıştır. Uzmanlar bu görüşmede tartışarak karar verdikleri için tam bir uyum sağlanmıştır. Görüşme yaklaşık 45 dakika sürmüştür. İnceleme formunda yer alan maddeler aşağıda sunulmuştur.

1. Ölçeklerin hedef grubu kimlerden oluşmaktadır?
2. Ölçeklerde
 - 2.a. Geliştirme için madde havuzu oluşturulmuş mudur?
 - 2.b. Uyarlama ise hangi çeviri yöntemi tercih edilmiştir, uyarlanan ölçeğin geliştirildiği kültür ve kaynak dil dağılımı nasıldır?
3. Ön uygulama gerçekleştirilmiş midir?
4. Kullanılan ölçek türleri nelerdir ve nasıl değişmektedir?
5. Ölçeklerde madde sayılarının dağılımı nasıl değişmektedir?
6. Ölçeklerin boyutlarının dağılımı ve isimlendirilmesi ve madde sayısı / boyut sayısı oranı nasıldır?
7. Ölçeklerde tercih edilen güvenilirlik kestirim yöntemleri nelerdir?
8. Ölçeklerde tercih edilen geçerlik kestirim yöntemleri nelerdir?
9. Ölçeklerde kesme puanı verilmiş midir?

Veri Analizi

Araştırmada tematik içerik analizi gerçekleştirilmiştir. Literatür taramasında kullanılan analizlerden birisi olan tematik içerik analizi, sıklıkla makalelerde tercih edilen bir yöntemdir (Abdul Razak ve diğerleri, 2023; Gezer, 2022; Şengül-Avşar ve Barış-Pekmezci, 2022). Tematik analizde araştırmacı temaları belirler ve çalışmalardan elde edilen ana fikir ve sonuçları tartışır (Cresswell, 2012). Çalışmada öncelikle çeşitli kaynaklara dayalı olarak temalar belirlenmiş ve ardından bu temalara yönelik kodlamalar gerçekleştirilmiştir. Araştırmada kodlamalar iki ayrı araştırmacı tarafından gerçekleştirilmiştir. Puanlayıcılar arası uyum Miles ve Huberman'a dayalı olarak hesaplanmış ve %94,44 olarak elde edilmiştir. Buna dayalı olarak puanlayıcılar arası güvenilirliğin yüksek olduğu söylenebilir. Uyuşmazlığa sebep olan kodlamalar iki kodlayıcı tarafından tekrar incelenmiş ve yeniden kodlanarak var olan uyumsuzluklar giderilmiştir.

Bulgular

Araştırmanın bulguları "Sınav Kaygısına Yönelik Kullanılan Ölçekleri İnceleme Formu" kapsamında yer alan maddeler doğrultusunda bu bölümde sunulmuştur.

Hedef Grup

Ölçeklerin geliştirildiği ya da uyarlandığı hedef gruba yönelik dağılım Tablo 2'de sunulmuştur.

Tablo 2

Ölçeklerin Hedef Grubunun Dağılımı

	Kategoriler	Frekans
Hedef Grup	Ebeveynler	1
	İlkokul ve Ortaokul Öğrencileri	1
	Ortaokul Öğrencileri	1
	Ortaokul ve Lise Öğrencileri	3
	Lise Öğrencileri	2
	Üniversite Öğrencileri	2

Tablo 2’de ölçeklerin hedef grubu incelendiğinde en çok ortaokul ve lise ile çalışıldığı görülmektedir. Ölçekler incelendiğinde ise bir çalışmada diğer çalışmalardan farklı olarak ebeveynler ile çalışıldığı görülmektedir (Baytemir & İlhan, 2018). Ayrıca ortaokul öğrencileri ile gerçekleştirilen bir çalışmanın 4. Sınıf ve üstü hedef grup için uygun olduğu belirtilmiş olup 8. sınıf liseye giriş sınavına hazırlanan öğrencilerden veri toplanılmıştır (Başol, 2017). Lise öğrencileri ile gerçekleştirilen bir çalışmanın ise üniversite sınavına hazırlanan son sınıf öğrencileri ile gerçekleştirildiği belirtilmiştir (Bozkurt et al., 2017).

Geliştirilen Ölçeklerde Madde Havuzu Oluşturma

Ölçek geliştirme çalışmaları (f = 5) incelendiğinde tüm çalışmalarda madde havuzu oluşturulduğu gözlemlenmiştir. Örneğin 18 maddenin bulunduğu “Ebeveyn Sınav Kaygısı Ölçeği” (Baytemir & İlhan, 2018) için 71 maddelik bir madde havuzu oluşturulurken, 22 maddelik “Durumluk Sınav Kaygısı Ölçeği”nin (Şahin, 2019) geliştirilmesi için 29 maddelik bir madde havuzu oluşturulmuştur.

Uyarlanan Ölçeklerde Kullanılan Çeviri Yöntemi, Kaynak Kültür ve Kaynak Dil

Ölçeklerin Türkçe’ye uyarlandığı çalışmalarda kullanılan çeviri yöntemlerine, kaynak kültürlerine ve kaynak dillere ilişkin dağılım Tablo 3’te verilmiştir.

Tablo 3

Uyarlanan Ölçeklerde Kullanılan Çeviri Yöntemi, Kaynak Kültür ve Kaynak Dil

	Kategoriler	Frekans
Çeviri Yöntemi	Geri Çeviri Yöntemi	4
	Tek Yönlü Çeviri	1
Kaynak Kültür	Amerikan	1
	Bahsedilmemiş	4
Kaynak Dil	İngilizce	5

Uyarlanan ölçeklerde kullanılan çeviri yöntemi incelendiğinde çalışmaların büyük bir çoğunluğunda (f = 4, %80.0) geri çeviri yönteminin kullanıldığı gözlemlenmiştir. Tek yönlü çeviri yöntemini kullanan tek bir çalışma söz konusudur ve bu çalışmada “Westside Sınav Kaygısı Ölçeği”nin ortaokul ve lise öğrencilerine yönelik geçerlik ve güvenilirlik çalışması yapılmıştır (Totan, 2018). Araştırmacı, Totan ve Yavuz (2009) tarafından gerçekleştirilen ilk uyarlama çalışmasında çevirinin alan bilgisi olan uzmanlar tarafından yapılması nedeniyle geri çeviri yapılmadığını ifade etmiştir. Uyarlanan ölçeklerin geliştirildiği kaynak kültürler incelendiğinde ise yine çalışmaların büyük bir çoğunluğunda (f = 4, %80.0) belirtilmediği gözlemlenmiştir. Kaynak kültüre ilişkin bilgi veren tek bir ölçek uyarlama çalışması bulunmaktadır. Bu çalışmada uyarlanan “Bilişsel Sınav

Kaygısı Ölçeği–Revize”nin Amerika’da geliştirildiği belirtilmiştir (Bozkurt ve diğerleri, 2017). Ölçek uyarlama çalışmalarında ele alınan ölçeklerin geliştirildiği dil incelendiğinde tüm ölçeklerin kaynak dilinin İngilizce olduğu görülmüştür.

Ön Uygulama

Sınav kaygısına yönelik ölçeklerin geliştirildiği ya da uyarlandığı çalışmalarda ön uygulama yapılma durumuna ilişkin dağılım Tablo 4’te verilmiştir.

Tablo 4

Ölçeklerin Ön Uygulama Yapılması Dağılımı

	Kategoriler	Frekans
Ön Uygulama	Yapılmış	7
	Bahsedilmemiş	3

Tablo 4 incelendiğinde ölçeklerin çoğunluğu için ön uygulamanın yapıldığı ($f = 7$, %70.0) görülmüştür. Doğrudan taranan çalışmaların içinde pilot çalışmaların yer aldığı durumlar olduğu gibi pilot çalışması başka bir çalışma içerisinde gerçekleştirilen bir çalışma da vardır. Totan (2018) tarafından yürütülen “Westside Sınav Kaygısı Ölçeği”nin geçerlik ve güvenilirlik çalışması kapsamında pilot çalışma yapıldığı söylenebilir. Çünkü halihazırda uyarlanmış bu ölçeğin uyarlama çalışmasında (Totan ve Yavuz, 2009) 20 öğrenci ile pilot çalışma yapıldığı belirtilmiştir.

Ölçek Türü

Araştırma kapsamında ele alınan ölçeklerin ölçek türü incelendiğinde, ölçeklerin biri hariç tamamında ($f = 9$; %90.0) Likert tipi ölçek kullanıldığı belirlenmiştir. Bir ölçekte ise kullanılan ölçek türüne değinilmemiştir. Likert tipi ölçeğin kullanıldığı bu ölçeklerdeki nokta sayısı dağılımı Tablo 5’te verilmiştir.

Tablo 5

Ölçeklerin Likert Tipi Dağılımı

	Kategoriler	Frekans
Likert Tipi	3’lü Likert ölçeği	1
	4’lü Likert ölçeği	3
	5’li Likert ölçeği	5

Tablo 5 incelendiğinde ise ölçeklerin en çok 5’li Likert ölçeğinde geliştirildiği görülmüştür. Bunu takip eden Likert türü ise 4’lü Likert ölçeğidir. Ölçeklerin hiçbirinde ölçek noktaları beşi geçmemektedir. Ölçeklerden yalnızca biri 3’lü Likert ölçeği olarak geliştirilmiştir (Tan, 2020) ve bu ölçek ortaokul ve lise öğrencilerine yöneliktir.

Madde Sayısı

Araştırma kapsamında ele alınan ölçeklere ilişkin madde sayısı dağılımı Tablo 6’da verilmiştir.

Tablo 6

Ölçeklerin Madde Sayısı Dağılımı

	Kategoriler	Frekans
Madde Sayısı	11-15 madde	2
	16-20 madde	3
	21-25 madde	3
	26-30 madde	2

Tablo 6’da ölçeklerin madde dağılımları incelendiğinde 30 maddeden fazla madde içeren ölçek olmadığı görülmüştür. En fazla madde sayısına sahip iki ölçek bulunmaktadır ve bu ölçeklerde 30 madde bulunmaktadır (Aydın ve Bulgan, 2017; Tan, 2020). En az maddeye sahip olan ölçek ise Totan (2018) tarafından geçerlik çalışması yürütülen “Westside Sınav Kaygısı Ölçeği”dir ve bu ölçek 11 maddeden oluşmaktadır. Bu ölçeği 15 madde ile Başol (2017) tarafından geliştirilen “AYDA Sınav Kaygısı Ölçeği” takip etmektedir.

Boyutların Dağılımı ve İsimlendirilmesi ve Madde Sayısı/Boyut Sayısı Oranı

Araştırma kapsamında ele alınan ölçeklerdeki boyutların dağılımı ve isimleri ile madde sayısı ve boyut oranı Tablo 7’de sunulmuştur.

Tablo 7

Boyutların Faktör Sayısı ve İsimlendirilmesinin Dağılımı

Ölçek Adı	Boyut Sayısı	Boyutların İsimleri (ilgili boyutta madde sayısı)
Ebeveynler için Sınav Kaygısı Ölçeği	2	Endişe (9), Psikolojik (9)
Online Sınavlarda Sınav Kaygısı Ölçeği	2	Fizyolojik ve Psikolojik Kaygı (12), Teknik Kaygı (8)
AYDA Sınav Kaygısı Ölçeği	2	Bilişsel ve Fizyolojik Kaygı (9), Aile ve Çevresel Kaygı (6)
Durumluk Sınav Kaygısı Ölçeği	3	Bilişsel (9), Psikososyal (5), Fizyolojik (8)
Bilişsel Sınav Kaygısı Ölçeği-Revize	1	Bilişsel (23)
Çocuklar ve Ergenler için Sınav Kaygısı Ölçeği	3 ve 4	Üç alt boyutlu model için; Biyolojik (10), Psikolojik (10), Sosyal (10) Dört alt boyutlu model için; Bilişsel (3), Psikolojik (10), Duygusal, (7) Davranışsal (10)
Çocuklarda Sınav Kaygısı Ölçeği	3	Düşünceler (13), Görev Dışı Davranışlar (8), Otonom Tepkiler (9)
Friedben Sınav Kaygısı Ölçeği	3	Bilişsel tıkanıklık, Gerginlik, Küçük Düşme (Madde sayıları belirtilmemiş)
Çok Boyutlu Sınav Kaygısı Ölçeği	4	Endişe (4), Bilişsel Müdahale (4), Gerginlik (4), Fizyolojik Göstergeler (4)
Westside Sınav Kaygısı Ölçeği	1	Sınav Kaygısı (11)

Tablo 7 incelendiğinde, ölçeklerin boyut sayılarının 1 ile 4 arasında değişmekte olduğu görülmüştür. Ölçekler boyut isimleri birbirlerinden farklılık göstermektedir. Boyutlarda en az 3 maddenin yer aldığı en çok ise tek boyutlu olarak belirlenen ölçekte 23 maddenin yer aldığı görülmektedir. Ayrıca bir ölçekte de alt boyutların madde sayıları sunulmamıştır. Ölçeklerde ele alınan bazı boyutların ortak isimlendirildiği de belirlenmiştir. Bulgularda dikkat çeken başka bir sonuç ise “Çocuklar ve Ergenler için Sınav Kaygısı Ölçeği”nde aynı veri üzerinde hem 3 boyutlu hem de 4 boyutlu yapı

incelenmiş ve her iki yapıya yönelik farklı isimlendirmeler kullanılmıştır.

Güvenirlilik Kestirim Yöntemi

Araştırma kapsamında ele alınan ölçeklerin güvenirliliğine kanıt sunmak amacıyla hesaplanan katsayıların dağılımı Tablo 8’de sunulmuştur.

Tablo 8

Güvenirlilik Katsayısı Türlerinin Dağılımı

		Kategoriler	Frekans
Güvenirlilik Katsayısı		Cronbach alfa	10
		Bileşik Güvenirlilik	1
		Spearman-Brown İki Yarı Korelasyonu Yöntemi	2
		Test-Tekrar Test	4

Tablo 8’de Cronbach alfa iç tutarlılık güvenirlilik katsayısının en çok hesaplandığı görülmektedir. İkinci olarak formun zaman içindeki tutarlılığını belirlemeye yarayan test-tekrar test (TTT) güvenirliliği kullanılmıştır. Bir çalışmada bileşik güvenirlilik katsayısı, Cronbach alfa ile birlikte hesaplanmıştır (Dikmen, 2023).

Tablo 9’da ölçeklerin genel ve alt boyutlarına ilişkin elde edilen güvenirlilik değerleri yer almaktadır. Tablo 9’da ölçeklerin Cronbach alfa katsayılarının 0,86 ila 0,94 arasında değiştiği görülmektedir. En yüksek güvenirlilik değerine bakıldığında “Online Sınavlar İçin Sınav Kaygısı Ölçeği” ve alt ölçeklerinden elde edildiği görülmektedir. Tüm ölçeklerde iç tutarlılık değerinin yüksek olduğu söylenebilir. TTT güvenirlilik değerinin 0,72 ile 0,81 arasında değiştiği görülmektedir. TTT güvenirlilik değeri ölçeklerde Cronbach alfa ile birlikte hesaplanmış ve genel olarak Cronbach alfa değerinden düşük çıkmıştır. Bu açıdan bakıldığında her iki güvenirlilik katsayısının hesaplandığı ölçeğin genelinde ve alt boyutlarda maddeler arası tutarlılığın zamana karşı tutarlılıktan daha yüksek olduğu belirlenmiştir. Cassidy ve Johnson (2002) tarafından geliştirilen Bilişsel Sınav Kaygısı Ölçeği’nin orijinal versiyonunda Cronbach alfa değeri 0.91 bulunmuş olup, ölçek için yüksek güvenirliliğe işaret etmektedir. Çocuklarda Sınav Kaygısı Ölçeğini (ÇSKÖ) geliştiren Wren ve Benson (2004) 0,92’lik bir Cronbach alfa değeri raporlamıştır ve bu da yüksek bir iç tutarlılık göstermektedir. Friedben Sınav Kaygısı Ölçeği, çeşitli çalışmalarda Cronbach alfa değerinin 0,85 ile 0,90 arasında değiştiğini göstermiştir ve bu da iyi bir güvenirliliğe işaret etmektedir (von der Embse ve diğerleri, 2014). Çok Boyutlu Sınav Kaygısı Ölçeğinin genel iç tutarlılık katsayısı 0,88 olarak elde edilmiştir (Putwain ve diğerleri, 2020). Westside Sınav Kaygısı Ölçeğinin geliştirildiği çalışmada ise Cronbach alfa değeri raporlanmamıştır (Driscoll, 2007).

Tablo 9

Ölçeklere İlişkin Genel ve Alt Boyutların Güvenirlik Katsayı Değeri

Ölçek Adı	Genel Güvenirlik Katsayısı	Alt Boyutlar için Güvenirlik Katsayıları
Ebeveynler için Sınav Kaygısı Ölçeği Online Sınavlarda Sınav Kaygısı Ölçeği	$\alpha = 0,91$ birinci çalışma için, $\alpha = 0,93$ ikinci çalışma için $\alpha = 0,94$	Endişe ($\alpha = 0,88$) ve Psikolojik boyut ($\alpha = 0,91$) birinci ve ikinci çalışma için Fizyolojik ve Psikolojik kaygı ($\alpha = 0,93$) Teknik kaygı ($\alpha = 0,90$); Fizyolojik ve Psikolojik kaygı (CR = 0,94), Teknik kaygı (CR = 0,90)
AYDA Sınav Kaygısı Ölçeği	$\alpha = 0,86$ $q = 0,797$; $r_{TTT} = 0,811$	Bilişsel ve Fizyolojik kaygı ($\alpha = 0,89$), Aile ve Çevre kaygısı ($\alpha = 0,70$)
Durumluk Sınav Kaygısı Ölçeği	$\alpha = 0,94$; $q = 0,96$; $r_{TTT} = 0,81$	Bilişsel ($\alpha = 0,93$), Psikososyal ($\alpha = 0,84$), Fizyolojik ($\alpha = 0,85$); Bilişsel ($q = 0,89$), Psikososyal ($q = 0,86$), Fizyolojik ($q = 0,84$); Bilişsel ($r_{TRT} = 0,74$), Psikososyal ($r_{TTT} = 0,80$), Fizyolojik ($r_{TTT} = 0,78$)
Bilişsel Sınav Kaygısı Ölçeği- Revize Edilmiş	$\alpha = 0,93$; $r_{TTT} = 0,93$	-
Çocuklar ve Ergenler için Sınav Kaygısı Ölçeği	$\alpha = 0,91$	Biyolojik ($\alpha = 0,83$), Psikolojik ($\alpha = 0,85$), Sosyal ($\alpha = 0,83$)
Çocuklarda Sınav Kaygısı Ölçeği	$\alpha = 0,88$	Düşünceler ($\alpha = 0,82$), Görev Dışı Davranışlar ($\alpha = 0,72$), Otonom Tepkiler ($\alpha = 0,75$)
Friedben Sınav Kaygısı Ölçeği	$\alpha = 0,88$	Bilişsel Tıkanıklık ($\alpha = 0,76$), Gerginlik ($\alpha = 0,81$), Küçük Düşürme ($\alpha = 0,78$)
Çok Boyutlu Sınav Kaygısı Ölçeği	$\alpha = 0,86$	Endişe ($\alpha = 0,84$), Bilişsel Müdahale ($\alpha = 0,82$), Gerginlik ($\alpha = 0,83$), Fizyolojik Göstergeler ($\alpha = 0,78$)
Westside Sınav Kaygısı Ölçeği	Ortaokul öğrencileri için $\alpha = 0,92$, Lise Öğrencileri için $\alpha = 0,91$; Ortaokul öğrencileri için $r_{TTT} = 0,72$, Lise Öğrencileri için $r_{TTT} = 0,74$	-

Not: α = Cronbach alfa güvenirlik katsayısı, q = Spearman-Brown iki yarı güvenirligi, r_{TTT} = Test-tekrar test güvenirlik katsayısı, CR = Bileşik güvenirlik katsayısı.

Geçerlik Kanıtı

Araştırma kapsamında ölçeklerin geçerliğine kanıt sağlamak amacıyla hesaplanan katsayıların dağılımı Tablo 10'da sunulmuştur.

Tablo 10

Geçerlik Kanıtlarının Dağılımı

Geçerlik	Kategoriler	Frekans
Yapı Geçerliği	Açımlayıcı Faktör Analizi	7
	Doğrulayıcı Faktör Analizi	9
Ölçüt Geçerliği	Yakınsak Geçerlik	6
	Iraksak Geçerlik	1
Kapsam Geçerliği		0

Tablo 10'da ölçeklerde en sık yapı geçerliğinin incelendiği, ikinci sırada ölçüt geçerliğinin incelendiği görülmektedir. Çalışmalarda kapsam geçerliği incelenmemiştir. Yapı geçerliğinin incelendiği çalışmalarda AFA ve DFA sonuçlarına ilişkin bulgular Tablo 11'de sunulmuştur. Ayrıca Tablo 12'de ölçüt geçerliğinin incelendiği çalışmalarda kullanılan ölçekler sunulmuştur.

Tablo 11

Ölçme Araçlarının AFA (açıklanan varyans) ve DFA (uyum indeksleri) Sonuçları

Ölçek Adı	Açıklanan Varyans	Kriterler
Ebeveynler için Sınav Kaygısı Ölçeği	%55,44	$\chi^2/sd=2,02$, RMSEA=,058, SRMR=,075, CFI=,98, NFI=,97, TLI=,98
Online Sınavlarda Sınav Kaygısı Ölçeği	%63,4	$\chi^2/sd=3,242$, RMSEA=,075, SRMR=,062, CFI=,89, GFI=,87, TLI=,87
AYDA Sınav Kaygısı Ölçeği	%53,73	$\chi^2/sd = 3,06$, RMSEA=,070, NNFI=,96, CFI=,97, GFI=,92, AGFI=,89
Durumluk Sınav Kaygısı Ölçeği	%59,21	$\chi^2/sd=1,72$, CFI=,96, NNFI=,96, IFI=,96, RMSEA=,05, SRMR=,05
Bilişsel Sınav Kaygısı Ölçeği-Revize	%38,8	CFI=,984, TLI=,982, RMSEA=,044, SRMR=,056
Çocuklar ve Ergenler için Sınav Kaygısı Ölçeği	Üç boyutlu model: %43,00 Dört boyutlu model: %47,64	-
Çocuklarda Sınav Kaygısı Ölçeği	-	$\chi^2=1414,72$, $sd=356,35$, $p=,00$, $\chi^2/df = 3,97$, RMSEA=,05, SRMR=,05, CFI=,97, GFI =,92, AGFI=,90
Friedben Sınav Kaygısı Ölçeği	-	$\chi^2=237,603$, $sd=95$, RMSEA=,05, CFI=,94, TLI=,92
Çok Boyutlu Sınav Kaygısı Ölçeği	-	$\chi^2= 669,27$, $sd=216$, RMSEA=,071, CFI=,87, GFI=,88, IFI=,88, SRMR=,073
Westside Sınav Kaygısı Ölçeği	%55,00	Ortaokul öğrencileri: $\chi^2= 137,96$, $sd=44$, $\chi^2/sd= 3,14$, CFI=,97, SRMR=,003 Lise öğrencileri: $\chi^2=128,33$, $sd=44$, $\chi^2/sd=2,92$, CFI=,96, SRMR=,003

Tablo 11 incelendiğinde altı ölçek için hem AFA hem de DFA analizi yapıldığı görülmüştür. Bir çalışmada yalnızca AFA, üç çalışmada ise DFA raporlanmıştır. Açıklanan varyans değerleri incelendiğinde en düşük açıklanan varyansın Bilişsel Sınav Kaygısı Ölçeği-Revize formunda (Bozkurt ve diğerleri, 2017) belirlendiği görülmektedir. Ancak ölçek tek boyutlu olduğundan elde edilen açıklanan varyans oranı yeterlidir. Tüm ölçeklerde DFA sonuçlarına göre elde edilen uyum indekslerinin genel olarak kabul edilebilir ve mükemmel uyum gösterdiği söylenebilir. Sonuç olarak araştırma kapsamında ele alınan ölçeklerin ilgili yapıyı temsil ettiği yorumu yapılabilir. Çok Boyutlu Sınav Kaygısı Ölçeği (Ergin ve Ekşi, 2022), Friedben Sınav Kaygısı Ölçeği (Akın ve diğerleri,

2013) ve Çocuklarda Sınav Kaygısı Ölçeği (Aydın ve Bulgan, 2017) için yapı geçerliliği DFA ile incelenirken, diğer ölçeklerde yapı geçerliliğini incelemek amacıyla AFA ve DFA kullanılmıştır.

Tablo 12

Ölçüt Geçerliliği Kanıtlarının Dağılımı

Ölçek Adı	Kriterler
Ebeveynler için Sınav Kaygısı Ölçeği	“Durum ve Süreklilik Kaygısı Ölçeği”
Online Sınavlarda Sınav Kaygısı Ölçeği	Yakınsak geçerlik: Faktör yükleri ve AVE, İraksak geçerlik: Fornell-Larcker kriteri ve Heterotrait-Monotrait
AYDA Sınav Kaygısı Ölçeği	“Spielberger Sınav Kaygısı Ölçeği” ve “Westside Sınav Kaygısı Ölçeği”
Durumluk Sınav Kaygısı Ölçeği	“Revize Edilmiş Sınav Kaygısı Ölçeği”
Bilişsel Sınav Kaygısı Ölçeği-Revize	“Test Kaygısı Envanteri”
Çocuklar ve Ergenler için Sınav Kaygısı Ölçeği	-
Çocuklarda Sınav Kaygısı Ölçeği	-
Friedben Sınav Kaygısı Ölçeği	-
Çok Boyutlu Sınav Kaygısı Ölçeği	“Revize Edilmiş Sınav Kaygısı Ölçeği”
Westside Sınav Kaygısı Ölçeği	-

Tablo 12, altı çalışmada ölçüt geçerliliğinin incelendiğini göstermektedir. Bu çalışmaların hepsinde hem yapı geçerliliği hem de ölçüt geçerliliği birlikte incelenmiştir. Sadece bir ölçek için hem yakınsak hem de iraksak geçerlik için farklı ölçütler kullanılmıştır (Dikmen, 2023). Diğer ölçekler için sadece yakınsak geçerlik analizleri yapılmıştır.

Kesme Puanı

Sadece bir ölçeğin kesme puanı vardır. AYDA Sınav Kaygısı Ölçeği 'nden alınan toplam puan 45 puanın altındaysa, endişe düzeyinde sınav kaygısı yoktur. 45-60 puan arası orta düzeyde sınav kaygısı, 60 puan üstü ise yüksek düzeyde sınav kaygısı olduğunu gösterir (Başol, 2017). Diğer dokuz ölçekte kesme puanı yoktur.

Tartışma

Bu çalışmada, sınav kaygısı yapısını ölçmek üzere Türk örnekleminde geliştirilen veya uyarlanan ölçekler çeşitli değişkenler bağlamında incelenmiştir. Çalışmanın ilk amacı, ölçeklerin uygulandığı grubun özelliklerini incelemektir. Sınav kaygısı ölçeklerinin hedef grupları, farklı yaş aralıklarını, eğitim düzeylerini ve bağlamları kapsayacak şekilde büyük ölçüde değişmektedir. Ölçekler genellikle ilkokuldan üniversiteye kadar değişen düzeylerdeki öğrencileri hedef almaktadır; çünkü bu öğrenciler sıklıkla akademik değerlendirmelere maruz kalmakta ve sınav kaygısı yaşayabilmektedirler (Cassady ve Johnson, 2002; Zeidner, 1998). Türkiye'de, eğitim sisteminin yapısı ve sınav odaklı yaklaşımı nedeniyle sınav kaygısı belirli dönemlerde yoğunlaşmaktadır. Özellikle lise giriş sınavlarının ortaokuldan sonra, üniversite giriş sınavlarının ise liseden sonra yapılması nedeniyle bu öğrenci gruplarının sınav kaygısının incelenmesi yoğunluk göstermektedir. Literatür incelendiğinde, sınav kaygısının en çok görüldüğü yaş dönemlerinin 11-14 yaşları olduğu görülmektedir. Genç nüfusta %40 olarak görülen sınav kaygısının tüm üniversite öğrencilerinde %20 ile %35 arasında değiştiği ifade edilmektedir. (McDonald, 2001; Trigueros ve diğerleri, 2020). Benzer şekilde Türkiye'de üniversite sonrası işe alım sınavları olması nedeniyle üniversite öğrencilerinden elde edilen verilerde de sınav kaygısı yapısı incelenmiştir.

Ölçeklerin hemen hepsi Likert tipi olup, beşli Likert ölçeği en çok tercih edilen ölçek olmuştur. Beşli Likert ölçeğinin basitliği ve kullanım kolaylığı vardır. Bu basitlik yanıt oranını ve toplanan verilerin kalitesini artırır. Türkiye bağlamında beşli Likert ölçeğinin kültürel olarak uygun ve etkili olduğu bulunmuştur. Türk öğrenciler eğitim değerlendirmelerinde Likert tipi ölçeklere alışkındır ve bu da onu sınav kaygısı gibi psikolojik yapıları ölçmek için tanıdık ve kabul görmüş bir araç haline getirir (Tezbaşaran, 1997).

Ölçeklerdeki madde sayıları en az 11, en çok 30 madde olmakla birlikte çoğunlukla 16-25 madde aralığındadır. Madde çeşitliliği, araştırmaya veya klinik ihtiyaca göre hem hızlı değerlendirmelere hem de derinlemesine incelemelere olanak tanır (Sarason, 1984; Spielberger ve diğerleri, 1980). Ölçeklerin boyut sayıları incelendiğinde çoğunlukla 3 boyutlu bir yapının benimsendiği görülmektedir. Ölçeklerin faktör yapılarının adları her birinde farklı olmakla birlikte psikolojik, fizyolojik, bilişsel, gerginlik, kaygı olarak adlandırılan boyutlar ortak adlarla birden fazla ölçekte ele alınmıştır. Sınav kaygısı ölçeklerinin çok boyutlu yapılara sahip olduğu ve genellikle bilişsel ve fizyolojik faktörleri içerdiği görülmektedir (Friedman ve Bendas-Jacob, 1997; Lowe ve Lee, 2008; Sarason ve Sarason, 1990; Wren ve Benson, 2004). Literatürde sınav kaygısı belirtilerinin bilişsel, fizyolojik, davranışsal ve duygusal belirtiler olmak üzere dört ana boyut altında sınıflandırıldığı görülmektedir (Yeşilyurt, 2007). Bu nedenle sınav kaygısı ölçeklerinin faktör yapısı belirlenirken ilgili tüm boyutlar dikkate alınmalıdır. Boyutlardaki madde sayılarına bakıldığında, ölçeklerden üçünün alt boyutlarda tutarlı sayıda maddeye sahip olduğu, diğer ölçeklerin alt boyutlarında madde sayılarının değiştiği görülmektedir. Madde oranları faktörlere göre değişmektedir; bazı ölçekler dengeli bir dağılıma sahipken, diğerleri belirli boyutları daha fazla vurgulamaktadır. Örneğin, Bilişsel Sınav Kaygısı Ölçeği (CTAS), Türkçe versiyonunda olduğu gibi (Bozkurt ve diğerleri, 2017), hem faktör yapısından hem de madde dağılımından anlaşılacağı üzere sınav kaygısının bilişsel yönlerini vurgulamaktadır. Ölçeklerdeki madde sayılarının dağılımı değişmektedir; bazı ölçekler sınav kaygısının farklı yönlerini kapsamlı bir şekilde yakalamak için daha fazla sayıda madde içermektedir (Cassady ve Johnson, 2002).

İncelenen tüm ölçek geliştirme çalışmalarında, maddelerin sınav kaygısı yapısını kapsamlı ve doğru bir şekilde ele aldığı ve hedef grupla ilgili olduğundan emin olmak için madde havuzları oluşturulmuştur. Bir madde havuzunun geliştirilmesi ve analizi, etkili sınav kaygısı ölçekleri oluşturmada kritik adımlardan biridir. Araştırmacılar, madde havuzu geliştirirken ölçülecek kapsamı, uzman görüşlerini ve kültürü göz önünde bulundurmalıdır. Böylelikle geçerliği yüksek sınav kaygısı ölçekleri geliştirerek eğitim uygulamalarına ve politikalarına doğru bilgi sağlamış olurlar. Bu nedenle pilot uygulamalar madde havuzu ve madde performansı için önemlidir. Pilot uygulamaların eksikliği, sınav kaygısı ölçeklerinin geçerliği ve güvenilirliği için önemli potansiyel riskler oluşturmaktadır. Araştırmaya dâhil edilen ölçek geliştirme çalışmalarının çoğunluğu, ölçeklerin ilk versiyonlarını test etmek, geri bildirimlere dayanarak ölçekleri iyileştirmek ve ölçeklerin sınav kaygısını ölçmeye uygunluğunu incelemek için pilot çalışmalar yürütmüş olsa da, çalışmaların bir kısmı pilot çalışmaların yürütülüp yürütülmediğine ilişkin bilgi içermemektedir. Pilot uygulamalar genellikle küçük, temsili örnekleri içerir ve ölçekteki değişiklikler için önemli geri bildirim sağlar (Cassady ve Johnson, 2002).

Dolayısıyla pilot uygulaması yapılmayan ölçeklerde, geçerlik testlerinin ön aşamaları göz ardı edileceğinden, hatalı sonuçlara ulaşma riski artmaktadır. Aynı şekilde bu aşamanın eksikliği güvenilirliği de etkilenmektedir ve pilot çalışmalarda genellikle güvenilirliğe ilişkin veriler de elde edilir. Bu nedenle pilot çalışma yapılmaması, bir ölçekte önemli bir metodolojik eksiklik olarak değerlendirilmeli ve ölçümlerin güvenilirliğini ve geçerliğini etkileyen temel bir faktör olarak

görülmemelidir.

Ölçek uyarlamada, özellikle farklı evrenleri içeren çalışmalarda, çeviri aşamasında kültürel uygunluğu ve dilsel eşdeğerliği sağlamak önemlidir (Furlan ve diğerleri, 2009). Bu amaçla tek yönlü çeviri, geri çeviri ve komite yaklaşımı yöntemleri yaygın olarak kullanılmaktadır (Brislin, 1970). Geri çeviri yöntemi, ölçeklerin bir tanesi hariç hepsinde kullanılmıştır. Ölçeklerin geliştirildiği kültürlerin dağılımına ilişkin olarak, Amerikan kültüründen uyarlanan bir ölçek olduğu (Bozkurt ve diğerleri, 2017) ve kaynak dil olarak geliştirilen tüm ölçeklerin dilinin İngilizce olduğu tespit edilmiştir.

Taranan sınav kaygısı ölçeklerinin tamamında güvenilirlik kestirim yöntemlerinden Cronbach alfa katsayısı raporlanmıştır. Buna ek olarak bazı ölçeklerde test-tekrar test güvenilirlik katsayısı ve Spearman-Brown iki yarı güvenilirliği yöntemi de raporlanmıştır. İncelenen tüm ölçekler için hesaplanan Cronbach alfa katsayıları, ölçeklerin yüksek iç tutarlılığa sahip olduğunu göstermektedir (genellikle 0,70'in üzerinde) (Nunnally ve Bernstein, 1994). Sınav kaygısı ölçekleri için test geliştiriciler genellikle 0,80 ve 0,90 gibi yüksek güvenilirlik değerlerine sahip ölçümler elde etmiştir (Zeidner, 1998). Ölçeklerin orijinal versiyonunun güvenilirlik puanları incelendiğinde, uyarlanmış ölçeklerin güvenilirlik puanları arasında genellikle büyük farklılıklar bulunmamaktadır. Bu durum, uyarlanmış ölçeklerin güvenilirlik değerleri konusunda endişe duyulmaması gerektiği sonucuna varmamıza yardımcı olmaktadır. Güvenirlik kestirim yöntemleri, ölçeklerin zaman içinde ve farklı evrenler arasında tutarlı ve kararlı sonuçlar üretmesini sağlamıştır (Nunnally ve Bernstein, 1994). Her ne kadar bileşik güvenilirlik, Cronbach alfa gibi, çok maddeli bir ölçeğin güvenilirliğini maddelerin faktör yüklerini ve hata varyanslarını dikkate alarak değerlendirirse de, incelenen ölçeklerden yalnızca biri için hesaplanmıştır. Türkiye'deki sınav kaygısı araştırmalarında, özellikle ölçeğin yapısını doğrulamak için doğrulayıcı faktör analizi (DFA) kullanan çalışmalarda, bileşik güvenilirlik giderek daha fazla raporlanmaktadır.

Araştırmaya konu olan ölçekler geçerlik kanıtları açısından incelendiğinde, ölçekler için en çok dikkate alınan geçerlik türünün yapı geçerliği olduğu belirlenmiştir. Yapı geçerliliği, bir değerlendirme aracının ölçülmesi hedeflenen yapıyı ne kadar iyi ölçtüğü ile ilgilidir (Haynes ve diğerleri, 1995). Taranan çalışmalarda DFA, genellikle açımlayıcı faktör analizi (AFA) ile birlikte yürütülmüştür. AFA ölçeğin yapısı hakkında ilk içgörülerini sağlarken, DFA faktör yapısını ve değişkenler arasındaki ilişkileri doğrulayarak sonuç olarak ölçüm aracının geçerliliğini sağlamaktadır (Marsh ve diğerleri., 2014). AFA genellikle daha fazla veri ve karmaşık analiz gerektirirken, DFA daha az veriyle daha hızlı sonuçlar sağlar. Ölçek uyarlama çalışmalarında, ölçek daha önce başka bir kültürde geliştirildiği ve faktör yapısı belirlendiği için genellikle AFA yapılmadan sadece DFA uygulandığı görülmektedir. Bazı araştırmacılar AFA'nın sağladığı keşfedici bilgileri gereksiz bularak doğrudan doğrulayıcı bir yaklaşım benimserken (Orçan, 2018), bazıları ise DFA ve AFA'nın birlikte kullanılmasının daha iyi sonuçlar verdiğini vurgulamaktadır (Marsh ve diğerleri, 2014). Araştırmacıların ölçeği, geliştirildiği kültüre uygunluğunu değerlendirmek için sadece DFA ile incelemeleri zaman ve kaynak tasarrufu açısından makul olabilir. Kapsam geçerliği, bir ölçme aracının ölçmeyi amaçladığı kavramı ne kadar iyi temsil ettiğini belirler. Kapsam geçerliğine ilişkin kanıt toplamak yapının bütüncül bir şekilde ele alındığını gösterir. İncelenen çalışmalarda kapsam geçerliğine dair bilgi verilmemiş olması geçerliğe ilişkin yeterli kanıt sunulmadığını göstermektedir. Çoğu araştırmacı geçerlik ve güvenilirlik testlerine odaklanırken, kapsam geçerliği gibi daha öznel ölçütleri raporlamayı göz ardı etmektedir. Araştırmacılar raporlamada yer vermedikleri için kapsam geçerliğini ne düzeyde dikkate aldıkları kestirilememektedir. Kapsam geçerliği sağlanmamış bir ölçek kullanılması durumunda geçersiz

ölçümler yanlış sonuçlara ve yanıltıcı verilere yol açabilir ve bu da yanlış yönlendirilmiş eğitim politikalarına ve müdahale programlarına yol açabilir. Ayrıca öğrencilerin sınav kaygısını etkileyen psikolojik ve sosyal faktörlerin yeterince anlaşılması, bu durumun yönetilmesi açısından olumsuz sonuçlar doğurabilmektedir. Bu nedenle geliştirilen ölçekler için kapsam geçerliğini sağladıklarına dair bilgi sunulması beklenmektedir.

Ölçek geliştirme ve uyarlama çalışmaları ölçüt geçerliği açısından incelendiğinde, yakınsak geçerlik, sınav kaygısı puanları ile genel kaygı veya performans kaygısı gibi benzer yapıların ölçümleri arasındaki korelasyonun incelenmesiyle değerlendirilmiştir. Yüksek korelasyon değerleri, sınav kaygısı ölçeğinin teorik olarak diğer kaygı biçimleriyle ilişkili bir yapıyı ölçtüğünü gösterir. İraksak geçerlik incelendiğinde ise sınav kaygısı ölçeğinin depresyon veya öz saygı gibi diğer yapıların ölçümleriyle ilişkili olmadığını göstermeyi amaçlar. Bu ilgisiz yapılarla düşük korelasyon değerleri, sınav kaygısı ölçeğinin bu yapılardan ayrı bir yapıyı ölçtüğünü gösterir (Campbell ve Fiske, 1959). Rosas ve Furlan (2017), sınav kaygısı ölçeklerinin amaçlanan yapıyı doğru bir şekilde ölçtüğünden emin olmak için yakınsak ve ıraksak ölçüt geçerliğini oluşturmanın önemini vurgular. Geçerlik süreci, ölçeğin sınav kaygısını değerlendirmede geçerliğini, ilgili yapılarla beklediği gibi ilişkili olduğunu ve ilgisiz yapılardan farklı olduğunu göstererek doğrulamayı içerir.

Ölçek uyarlama ve geliştirme çalışmalarında, yalnızca bir ölçeğin kesme puanına sahip olduğu görülmüştür. Literatür incelendiğinde, sınav kaygısı ölçeklerinde kesme noktalarının varlığı çalışmalar arasında farklılık göstermektedir. Kesme noktaları bulunan ölçekler bunu genellikle klinik olarak önemli düzeyde sınav kaygısı olan bireylerin belirlenmesini kolaylaştırmak için yapar. Örneğin, Sınav Kaygısı Envanteri; düşük, orta ve yüksek kaygı düzeylerini kategorize eden kesme noktalarına sahiptir (Spielberger, 1980). Sınav kaygısı ölçeklerindeki kesme noktaları, duyarlılık ve özgüllük arasındaki dengeyi göz önünde bulundurarak sınav kaygısı düzeylerini etkili bir şekilde değerlendirmeyi sağlar. Böylece vakaları gözden kaçırmadan veya yanlış sınıflandırmadan doğru tanımlama sağlanmaktadır. Öte yandan, bazı ölçeklerde kesme noktaları bulunmamaktadır, bunun yerine sınav kaygısının sürekli bir ölçümünü sağlamaya odaklanır. Bu ölçekler, sınav kaygısının ayrı kategoriler yerine bir spektrum olarak daha iyi anlaşıldığını savunmaktadır (Zeidner, 1998). Bu yaklaşım sayesinde bireylerin kaygı profillerinin anlaşılması ve ihtiyaçlarına göre kişiselleştirilmiş müdahale planları oluşturulması kolaylaşır. Kültürel farklılıklar nedeniyle kesme puanları farklı örnek gruplarında her zaman geçerli olmayabilir. Bu nedenle, kesme puanı olmayan değerlendirmeler farklı demografik gruplarda daha uyumlu ve esnek sonuçlar sağlayabilir. Kesme puanı olan ölçümlerde sürekli veriler kategorik verilere dönüştürüldüğünde boylamsal olarak verinin ayrıntılı şekilde yorumlanması güçleşir. Diğer yandan kesme puanı olmayan ölçümlerde, zaman içindeki değişiklikleri daha iyi gözlemlemek mümkündür. Örneğin, bir bireyin sahip olduğu kaygı düzeyindeki artış veya azalış, puan dağılımında daha kolay gözlemlenebilir ve bu da dinamik değişim sürecini değerlendirmede bir avantaj sağlamaktadır. Ölçeklerdeki kesme noktaları, kaygının şiddetini belirleyen ve belirlenen kaygı düzeylerine göre uygun müdahaleleri yönlendiren klinik veya pratik uygulamaları kolaylaştırmaktadır. Fakat, kesme noktasına sahip olmayan ölçekler daha kapsamlı bir değerlendirme için sürekli bir ölçme yaklaşımını sürdürür.

Farklı yaş gruplarındaki öğrencilerde sınav kaygısını etkileyen değişkenleri açıklamak için karşılaştırmalı çalışmalar yürütülürken aynı ölçeğin kullanılması gerekebilir. Bu doğrultuda ölçeklerin çeşitli gruplarda kullanıma uygunluğunu tespit etmek için, farklı sosyo-ekonomik ve demografik özelliklere sahip katılımcılarla daha fazla analiz yapılması faydalı olacaktır. Örneğin, ölçekler için lise ve üniversite düzeylerinden farklı olarak ilköğretim düzeyini de kapsayacak şekilde geniş bir yaş aralığında geçerlik çalışmaları gerçekleştirilebilir. Çünkü, sınav kaygısının erken

yaşlarda belirlenmesi, bu kaygının zaman içindeki değişimi hakkında önemli bilgiler sağlayabilir ve böylece eğitim süreçlerinde önleyici tedbirlerin uygulanmasını kolaylaştırabilir. Ayrıca araştırmacılar demografik özellikler, akademik geçmiş ve sosyal destek gibi faktörleri göz önünde bulundurarak daha kapsamlı ölçekler geliştirebilirler. Geliştirilmesi önerilen bu ölçeklerde sınav kaygısının çok boyutlu doğasının göz önünde bulundurulması yapının kapsamlı şekilde değerlendirilmesini sağlayacaktır. Tematik içerik analizinin kullanıldığı bu çalışmanın doğası gereği, değerlendirme ölçütleri araştırmacılar tarafından belirlendiğinden öznellik içerebilir. Gelecek araştırmalarda, literatürdeki başka ölçeklerin psikometrik özellikleri bu araştırmadaki değerlendirme ölçütleri veya bu ölçütlere ek farklı ölçütler kullanılarak incelenebilir.

Etik Kurul Onayı: Bu çalışma insan/hayvan araştırmaları kapsamına girmediği ve çalışma grubunu birincil çalışmalar oluşturduğu için etik kurul kapsamında değildir.

Araştırmacıların Katkı Oranı: Çalışmaya yazarlar eşit oranda katkı sunmuştur.

Çatışma Beyanı: Yazarlar arasında herhangi bir çıkar çatışması yoktur.

References

* Primary studies on the scales included in the research

Abdul Razak, N., Ma'amor, H., Ahmad, M. F., & Mohd Rashid, N. (2023). Thematic analysis to measure social anxiety among students. *Information Management and Business Review*, 15(4(SI)I), 465-471. [https://doi.org/10.22610/imbr.v15i4\(SI\)I.3621](https://doi.org/10.22610/imbr.v15i4(SI)I.3621)

*Akın, A., Akın, U., Sarıçam, H., Aşut, S., Arslan, S., Demirci, İ., Toprak, H., & Çardak, M. (2013). *Friedben Sınav Kaygısı Ölçeği Türkçe Formu'nun geçerlik ve güvenilirliği* [Validity and reliability of the Turkish form of the Friedben Exam Anxiety Scale] [Conference presentation]. International Conference on Innovation and Challenges in Education, Kütahya, Türkiye. <https://doi.org/10.13140/RG.2.2.21471.15528>

*Aydın, U., & Bulgan, G. (2017). Çocuklarda Sınav Kaygısı Ölçeği'nin Türkçe uyarlaması [Adaptation of Children's Test Anxiety Scale to Turkish]. *Elementary Education Online*, 16(2), 887-887. <https://doi.org/10.17051/ilkonline.2017.304742>

*Başol, G. (2017). AYDA Sınav Kaygısı Ölçeği: Geçerlik ve güvenilirlik çalışması [IDA Test Anxiety Scale: Validity and reliability study]. *The Journal of International Educational Sciences*, 4(13), 173-193. <https://doi.org/10.16991/inesjournal.1506>

*Baytemir, K., & İlhan, T. (2018). Development of the exam anxiety scale for parents: A validity and reliability study. *Electronic Journal of Research in Education Psychology*, 16(44), 223-241. <https://doi.org/10.25115/ejrep.v16i44.1945>

*Bozkurt, S., Ekitli, G. B., Thomas, C. L., & Cassady, J. C. (2017). Validation of the Turkish version of the Cognitive Test Anxiety Scale-Revised. *Sage Open*, 7, 1-9. <https://doi.org/10.1177/2158244016669549>

Brady, S. T., Hard, B. M., & Gross, J. J. (2018). Reappraising test anxiety increases academic performance of first-year college students. *Journal of Educational Psychology*, 110(3), 395-406. <https://doi.org/10.1037/edu0000219>

- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216. <https://doi.org/10.1177/135910457000100301>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105. <https://doi.org/10.1037/h0046016>
- Cassady, J., & Johnson, R. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270-295. <https://doi.org/10.1006/ceps.2001.109>
- Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology*, 97(2), 268-274. <https://doi.org/10.1037/0022-0663.97.2.268>
- Creswell, J. W. (2012). *Educational research planning, conducting, and evaluating quantitative and qualitative research*. Pearson.
- Culler, R. E., & Holahan, C. J. (1980). Test anxiety and academic performance: The effects of study-related behaviors. *Journal of Educational Psychology*, 72(1), 16-20. <https://doi.org/10.1037/0022-0663.72.1.16>
- Çakmak, A., Şahin, H., & Akıncı Demirbaş, E. (2017). 7. ve 8. sınıf ortaokul öğrencilerinin sınav kaygısı ve benlik saygısı arasındaki ilişkinin incelenmesi [The analysis of relationship between test anxiety and self-esteem in the case of 7th and 8th grade students]. *E-Kafkas Journal of Educational Research*, 4(2), 1-9. <https://doi.org/10.30900/kafkasegt.315182>
- *Dikmen, M. (2023). Test anxiety in online exams: Scale development and validity. *Current Psychology*, 42, 30210-30222. <https://doi.org/10.1007/s12144-022-04072-0>
- Doğan, U. (2020). Lise ve üniversite öğrencilerinde sınav kaygısının karşılaştırılması [Comparison of test anxiety in high school and university students]. *Muğla Sıtkı Koçman Üniversitesi Eğitim Fakültesi Dergisi*, 7(1), 35-42. <https://doi.org/10.21666/muefd.637421>
- Driscoll, R. (2007). *Westside test anxiety scale validation*. ERIC Digest, ED495968.
- *Ergin, A., & Ekşi, H. (2022). *Adaptation of the Multidimensional Exam Anxiety Scale to Turkish: A validity and reliability study* [Conference presentation]. 3. International scientific research and innovation congress, İstanbul, Türkiye. <https://toad.halileksi.net/wp-content/uploads/2022/10/cok-boyutlu-sinav-kaygisi-olcegi-21774-toad.pdf>
- Friedman, I. A., & Bendas-Jacob, O. (1997). Measuring perceived test anxiety in adolescents: A self-report scale. *Educational and Psychological Measurement*, 57, 1035-1045.
- Furlan, L. A., Cassady, J., & Pérez, E. (2009). Adapting the cognitive test anxiety scale for use with Argentinean university students. *International Journal of Testing*, 9(1), 3-19. <https://doi.org/10.1080/15305050902733448>
- Gezer, M. (2022). The thematic content analysis of the scales used in citizenship education. *International Journal of Assessment Tools in Education*, 9(1), 233-253. <https://doi.org/10.21449/ijate.884855>
- Gökkaya, F. (2019). The relationship between the test anxiety level of sixth grade students and the trait anxiety and cognitive distortions of mothers. *Journal of Social, Humanities and Administrative Sciences*, 2(9), 621-634.
- Güngör, C. & Saltürk, A. (2022). "Gizli buzlanma" olarak sınav kaygısı: üniversite öğrencilerinin metaforik ifadelerinin değerlendirilmesi [Exam anxiety as "hidden icing": evaluating university students' metaphoric expressions]. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 22(4), 1397-1418 <https://dx.doi.org/10.17240/aibuefd.2022.22.74506-867417>

- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247. <https://doi.org/10.1037/1040-3590.7.3.238>
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58(1), 47-77. <https://doi.org/10.3102/00346543058001047>
- Kapıkıran, Ş. (2020). The mediator role of academic resilience in the relationship between parental academic success pressure and support and test anxiety in adolescents. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 48, 409-430. <https://doi.org/10.9779/pauefd.521230>
- Kırkık, M., & Balkıs, M. (2022). Ergenlerde üstbiliş ve sınav kaygısı arasındaki ilişkide ruminasyon ve bilişsel esnekliğin aracılık rolü [The mediating role of rumination and cognitive flexibility in the relationship between metacognition and test anxiety in adolescents]. *Muğla Sıtkı Koçman Üniversitesi Eğitim Fakültesi Dergisi*, 9(1), 61-77. <https://doi.org/10.21666/muefd.895400>
- 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. <https://doi.org/10.2466/pr0.1967.20.3.975>
- Lowe, P. A., & Lee, S. W. (2008). Factor structure of the Test Anxiety Inventory for Children and Adolescents (TAICA) scores across gender among students in elementary and secondary school settings. *Journal of Psychoeducational Assessment*, 26, 231-246. <https://doi.org/10.1177/0734282907303773>
- Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *The Journal of Abnormal and Social Psychology*, 47(2), 166-173. <https://doi.org/10.1037/h0062855>
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10(1), 85-110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Matthews, G., Hillyard, E. J., & Campbell, S. E. (1999). Metacognition and maladaptive coping as components of test anxiety. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, 6(2), 111-125. [https://doi.org/10.1002/\(SICI\)1099-0879\(199905\)6:2<111::AID-CPP192>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-0879(199905)6:2<111::AID-CPP192>3.0.CO;2-4)
- McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, 21, 89-101.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Oetting, E. R. (1966). Examination anxiety: Prediction, physiological response and relation to scholastic performance. *Journal of Counseling Psychology*, 13(2), 224-227. <https://doi.org/10.1037/h0023407>
- Orçan, F. (2018). Exploratory and confirmatory factor analysis: Which one to use first? *Journal of Measurement and Evaluation in Education and Psychology*, 9(4), 414-421. <https://doi.org/10.21031/epod.394323>
- Özcan, N., & Arslan, R. (2018). Ergen annelerine uygulanan ebeveyn stresini yönetme programı'nın etkililiği [The effectiveness of parent stress management program to mothers with adolescent child]. *Sakarya University Journal of Education*, 8(2), 40-59. <https://doi.org/10.19126/suje.349339>
- Putwain, D. W., von der Embse, N. P., Rainbird, E. C. & West, G. (2020). The development and validation of a new Multidimensional Test Anxiety Scale (MTAS). *European Journal of Psychological Assessment*, 37(3), 236-246. <https://doi.org/10.1027/1015-5759/a000604>

- Rosas, J. S., & Furlan, L. A. (2017). Achievement emotions and achievement goals in support of the convergent, divergent and criterion validity of the Spanish-Cognitive Test Anxiety Scale. *International Journal of Educational Psychology*, 6(1), 67-92. <https://doi.org/10.17583/ijep.2017.2268>
- Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology*, 46(4), 929-938. <https://doi.org/10.1037/0022-3514.46.4.929>
- Sarason, I. G., & Sarason, B. R. (1990). *Test anxiety*. In *handbook of social and evaluation anxiety* (pp. 475-495). Springer US. https://doi.org/10.1007/978-1-4899-2504-6_16
- Spielberger, C. D. (1980). *Test anxiety inventory*. Consulting Psychologists.
- *Şahin, A. (2019). State Test Anxiety Scale (STAS): Validity and reliability study. *Trakya Eğitim Dergisi*, 9(1), 78-90. <https://doi.org/10.24315/tred.450423>
- Şengül Avşar, A., & Barış Pekmezci, F. (2022). Examination of motivation scales: Is the purpose academic promotion or the need to measure psychological constructs?. *Psycho-Educational Research Reviews*, 11(3), 774-791. https://doi.org/10.52963/PERR_Biruni_V11.N3.19
- *Tan, C. (2020). *Çocuk ve ergenler için sınav kaygısı ölçeğinin biyopsikososyal modelle geliştirilmesi ve psikometrik özelliklerinin incelenmesi [Development of Test Anxiety Scale for Children and Adolescents using biopsychosocial model and examination of its psychometric properties]* [Unpublished Master's thesis]. Fatih Sultan Mehmet Vakıf Üniversitesi.
- Terzi, B., Daş-Geçim, G. Y., & Topuz, İ. (2019). Hemşirelik birinci sınıf öğrencilerinin beceri uygulama sınavındaki sınav kaygısı ve problem çözme becerileri arasındaki ilişkinin incelenmesi [Investigation of the relationship between test anxiety and problem solving skills of first year nursing students in skills practice exam]. *Celal Bayar Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, 6(4), 289-296. <https://doi.org/10.34087/cbusbed.604108>
- Tezbaşaran, A. A. (1997). *Likert tipi ölçek geliştirme kılavuzu [Likert type scale development guide]*. Türk Psikologlar Derneği.
- *Totan, T. (2018). Ortaokul ve lise öğrencilerinde sınav kaygısının değerlendirilmesi: Westside Sınav Kaygısı Ölçeği [Investigation of test exam on middle and high school students: The Westside Test Anxiety Scale]. *Batı Anadolu Eğitim Bilimleri Dergisi*, 9(2), 143-155. <https://dergipark.org.tr/tr/download/article-file/580489>
- Totan, T., & Yavuz, Y. (2009). Westside sınav kaygısı ölçeğinin Türkçe formunun geçerlik ve güvenilirlik çalışması [The validity and reliability study of the Turkish version of Westside Test Anxiety Scale]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 9(17), 95-109.
- Trigueros, R., Padilla, A. M., Aguilar-Parra, J. M., Rocamora, P., Morales-Gázquez, M. J., & LópezLiria, R. (2020). The influence of emotional intelligence on resilience, test anxiety, academic stress and the Mediterranean diet. A study with university students. *International Journal of Environmental Research and Public Health*, 17(6), 2071. <https://doi.org/10.3390/ijerph17062071>
- von der Embse, N. P., Mata, A. D., Segool, N., & Scott, E.-C. (2014). Latent profile analyses of test anxiety: a pilot study. *Journal of Psychoeducational Assessment*, 32(2), 165-172. <https://doi.org/10.1177/0734282913504541>
- Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, 76(2), 92-104. <https://doi.org/10.1037/h0031332>
- Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: scale development and internal construct validation. *Anxiety, Stress & Coping*, 17(3), 227-240. <https://doi.org/10.1080/10615800412331292606>

- Wong, S. S. (2008). The relations of cognitive triad, dysfunctional attitudes, automatic thoughts, and irrational beliefs with test anxiety. *Current Psychology*, 27(3), 177-191. <https://doi.org/10.1007/s12144-008-9033-y>
- Yeşilyurt, F. (2007). *ÖSS ve OKS'de sınav kaygısı ve baş etme yolları*. Remzi.
- Zeidner, M. (1998). *Test anxiety: The state of the art*. Plenum.