



Comparison of the Psychometric Properties of Extended Likert-Type Scales and Classic Likert-Type Scales in Psychological Tests*

•Seda Gözlügöl¹, •Seher Yalçın²

¹ Ministry of Family and Social Services, Gaziantep, Türkiye

² Director of the Center for Measurement and Evaluation Research and Applications, Ankara University, Ankara, Türkiye

ABSTRACT

This study aims to compare the psychometric properties of classic Likert-type scales with those of the newly developed extended alternative Likert-type scales in the literature. Specifically, it compares total scores across different scale administration conditions and examines the psychometric properties of the extended-alternative form for child and adolescent groups. Designed as an explanatory mixed-methods design, the study collected and analyzed quantitative data and subsequently collected and analyzed qualitative data to explain the quantitative findings in greater detail. Purposeful sampling and convenience sampling methods were used in the study. The study group consisted of elementary and middle school students from Gaziantep and Kilis provinces during the 2023-2024 academic year. A total of 892 students participated in quantitative data collection and 20 participated in qualitative interviews. The 'Children's Test Anxiety Scale (CTAS)' was used in the study. No significant difference was found between the validity and reliability evidence for the forms. The analysis revealed that respondents' total scores differed significantly based on their scale-taking status, and the difference between forms within the same group was statistically significant. The expanded-alternative form enables students to interpret items more consistently and appropriately, although some items in both forms are answered through different cognitive processes.

Keywords: likert-type scale, extended-alternative likert-type scale, reliability estimates, validity evidence, thinking aloud.

* This article is based on Seda Gözlügöl's master's thesis titled "A Comparison of the Psychometric Properties of Extended Likert-Type Scales and Classical Likert-Type Scales in Psychological Tests," which she completed in 2024 at the Institute of Educational Sciences, Ankara University, under the supervision of Prof. Dr. Seher Yalçın.

1 Corresponding Author: Teacher, E-mail: sedaakrldg@gmail.com **ORCID ID:** <https://orcid.org/0009-0003-1405-7550>
ROR ID: <https://ror.org/04n9e8n41>

2 Prof. Dr., E-mail: yalcins@ankara.edu.tr, **ORCID ID:** <https://orcid.org/0000-0003-0177-6727>, **ROR ID:** <https://ror.org/01wntqw50>

Received Date: 03.11.2025 **Accepted Date:** 02.06.2026 **Publication Date:** 06.08.2026

Citation: Gözlügöl, S., & Yalçın, S. (2026). Comparison of the psychometric properties of extended likert-type scales and classic likert-type scales in psychological test. *Ankara University Journal of Faculty of Educational Sciences*, 59(2), 591-638. <https://doi.org/10.30964/auebfd.1816495>

All ethical declarations related to this article are provided on the declarations page (Page: 632).

Psychological tests are measurement tools developed to assess people's behaviors, performance, attitudes, abilities, in short, all psychological characteristics, and to reveal their current state. They are standardized tools that can be used by qualified professionals (Cansız, 2007; Kutanis & Elçi, 2006). When used to make accurate decisions about individuals, for diagnosis, selection, and placement, as well as to reveal individual differences or the changes and development within the same person over time, the test must accurately reveal the situation of the individual being tested. This depends on the suitability of the measurement tool used for the purpose, the environment in which the measurement is made, the expertise of the person performing the measurement, and the psychometric characteristics of the measurement tools, such as reliability, validity, usability (Börkan et al., 2017; Cohen & Swerdlik, 2018; Şahin & Gülleroğlu, 2013; Tülübaş, 2009).

In classic Likert-type scales, since the respondent's statements reflect their opinions, attitudes, and feelings individually, the scores obtained from the statements are summed up to determine a total score from the responses given by individuals, and scaling is performed with rating totals (Gökdemir & Yılmaz, 2023; Şahin & Gülleroğlu, 2013; Turan et al., 2015). Classical Likert-type scales are used in many fields due to their high psychometric properties, the comparability of the total scores obtained from the scale with the total scores obtained from the sub-dimensions of the scale, the collection of data with standardized measurements, and the possibility of conducting validity-reliability analyses by making repeated measurements (Gökdemir & Yılmaz, 2023). Despite their advantages, classical Likert-type scales have limitations. Although classical Likert-type scales have recently been frequently used in the assessment, diagnosis, and treatment of children (Royeen, 1985; Uyumaz & Çokluk, 2016), the analysis of the psychometric properties of classical Likert-type scales mostly involves the evaluation of adult groups. the younger age group in early childhood has been included in studies to a limited extent, and the information provided to improve the psychometric properties of classic Likert-type scales used in child groups is limited. This situation limits the use of classic Likert-type scales and the assessment of their psychometric properties in younger age groups. Despite the frequent use of classic Likert-type scales in the assessment, diagnosis, and treatment of children (Royeen, 1985; Uyumaz & Çokluk, 2016), the analysis of the psychometric properties of classic Likert-type scales has mostly been conducted using adult groups. This situation casts doubt on the use of classic Likert-type scales in younger age groups and their psychometric properties (Alan, 2019; Borgers & Hox, 2001). The degree of uncertainty in the agreement on the items and response options in the classic Likert-type scale, the fact that the items are not understood in the same way by everyone, the fact that the items in the scale cause different interpretations by different individuals, the ambiguity of what the item means (Frequency? Degree?) is ambiguous, the inability to fully reflect the respondent's views and attitudes, and the fact that the amount corresponding to the ratings in the response category varies from person to person are limitations of the Likert-type scale and cause its psychometric properties to decline (Başaran, 1997; Joshi et al., 2015; Kam et al., 2024). One

disadvantage of classic Likert-type scales is that individuals who obtain the same total score with different response statements are considered to have the same profiles. Another disadvantage of the classic Likert-type scale is that it cannot reveal how much individuals differ from one another (Alan, 2019). In one of the studies conducted in recent years on the validity and reliability of classic Likert-type measurements (Zhang & Savalei, 2016), a new dimension was added to classic Likert-type scales. The researchers stated that responding to items in classic Likert-type scales with “agree-sometimes-disagree” is an ambiguous expression of the frequency or degree of behavior and therefore does not provide very specific results for the respondent. The researchers proposed an extended-alternative scale format, called the extended form, in which a complete sentence replaces each response option, and ambiguous expressions in the item are eliminated or reduced.

Extended-alternative Likert-type scales were first proposed by Hills and Arglye in their 2002 study titled “The Oxford Happiness Questionnaire: A compact scale for the measurement of psychological well-being.” Based on the Oxford happiness scale, they developed an extended Likert-type scale following the principle of presenting each rating statement as a sentence, rather than asking respondents to mark rating statements arranged on the scale from ‘agree’ to “disagree” to “strongly disagree.”

Zhang and Savalei (2016) argued in their study that extended-alternative Likert-type scales have higher psychometric properties and better explain individual profiles. The extended-alternative form reduces respondent confusion and helps them answer more carefully by enabling them to read the items more attentively and to understand the differences between the options more easily. In contrast, the items in the classic Likert form scales are ambiguous regarding the extent to which the respondent agrees or disagrees and are not open to evaluation (Kam et al., 2024). On a classic Likert-type scale, respondents must select either the “disagree” or ‘agree’ option, whereas expanded-alternative form items do not require respondents to agree with any statement, thereby eliminating the “disagree-agree” options and theoretically reducing consent bias. Respondents will not experience confusion when answering the items on the scale, as there is no transition from positively to negatively worded items and the response options for each item are specified in the sentence. This reduces respondent bias. While items in the classic Likert-type scale have a positive/negative direction, items in the expanded-alternative form have no direction; both positive and negative versions of each item are presented as response options (Zhang et al., 2019; Zhang et al., 2023).

Responding to items on an extended-alternative Likert-type scale requires more reading effort than responding to items on a classic Likert-type scale. Due to the length of the items on the extended-alternative Likert-type scale, respondents may experience fatigue, boredom, and inattention, leading to unwanted responses. Therefore, respondents must exercise greater care when reading the items. This situation may compromise the scale's psychometric properties. The development, application to a group, and evaluation of the extended-alternative Likert-type scale are

not economical in terms of time and cost because it takes longer for respondents to complete the items on the scale than the classic Likert-type form (Zhang et al., 2019). By converting the items of the classic Likert-type scale into items in the extended-alternative Likert-type form, it is thought that both researchers and respondents will understand the intended meaning of the response options more clearly, that the possibility of different interpretations of the statements by individuals will be reduced, and that the dimension that the scale aims to measure will be revealed more validly. Therefore, Zhang and Savalei (2016) recommend using alternative forms instead of classical Likert-type scales.

Research on extended-alternative Likert-type scales in the literature appears to be limited to studies conducted by the aforementioned researchers and to studies involving adult groups; no study on extended-alternative Likert-type scales has been found in Türkiye. Therefore, it examines and compares the psychometric properties of classic and extended-alternative Likert-type scales and clearly presents individuals' profiles using alternative Likert-type scales. In this study, it was deemed necessary to convert a Likert-type psychological test into extended-alternative Likert-type scales, to compare them in terms of psychometric properties, and to examine differences in validity and reliability separately for the child and adolescent groups, as scale comprehensibility and interest pose greater challenges in the child group. The aim was to compare the psychometric properties of the test anxiety scale for children, originally rated on a classic Likert scale, with those obtained after conversion to extended-alternative Likert scales, and to examine the effect of age group (child versus adolescent) on the validity and reliability of the extended form. Additionally, the study aimed to examine whether the scale total score and student profiles (those with low, medium, and high levels of the trait measured by the subscales) differed between the expanded-alternative and classic Likert-type scales.

Sub-Objectives

1. What are the validity and reliability evidence obtained from extended-alternative and classic Likert-type scales applied to different samples?

- a) What is the difference between the item factor loadings of the scales?
- b) What is the difference between the fit indices?
- c) How do reliability coefficients differ?

2.a) Do the validity and reliability findings for the scores of students who completed only the classic Likert-type scale differ between the child and adolescent groups?

b) Do the validity and reliability findings for the scores of students who completed only the expanded-alternative Likert-type scale differ between the child and adolescent groups?

3. Do the total mean scores obtained from the extended-alternative Likert-type scales differ from the total mean scores obtained from the classic Likert-type scales among students in the same group?

4. How do students interpret the statements in the classic Likert form and the extended-alternative form?

Method

Research Model

The study used an explanatory design with mixed models to compare the psychometric properties of the test anxiety scale for children, which was originally rated using a classic Likert-type scale, by converting it into extended-alternative Likert-type scales. This design is two-stage. First, quantitative data were collected and analyzed. Then, qualitative data were collected and analyzed to explain and examine the quantitative data in detail (Creswell & Plano-Clark, 2007). In this context, the test anxiety scale was first administered to students and analyzed, followed by individual interviews with 20 students using the think-aloud approach to examine how they interpreted the different response options.

Study Group

The Test Anxiety Scale for Children is an assessment tool that provides information about the test anxiety levels of students aged 8-12, and it is used both abroad and in our country. In this study, since it does not aim to generalize to the entire population, students were selected from schools affiliated with the Ministry of National Education in the provinces of Gaziantep and Kilis using non-random sampling methods. Purposeful sampling was used to conduct in-depth research and to select situations appropriate to the purpose of the study, and convenience sampling was used because of limitations of time, funding, labor, and other practical constraints. For the quantitative phase of the study, the Test Anxiety Scale for Children was administered to 892 students in grades 3, 4, 5, and 6 who volunteered to participate in the study, from schools affiliated with the Ministry of National Education in the provinces of Gaziantep and Kilis, during the first semester of the 2023-2024 academic year. Subsequently, during the qualitative phase of the study, individual interviews were conducted with 20 students in grades 3, 4, 5, and 6 who volunteered and possessed strong reading comprehension skills; the sample exhibited a balanced demographic distribution. A total of 912 students participated in the study, including 445 girls and 467 boys. Of the participating students, 598 were in the child group and 313 were in the adolescent group.

Data Collection Tools

In this study, the Children's Test Anxiety Scale was used as the data collection tool. Permission to use the scale in this study was obtained from the authors who developed it. Information about the scale is provided below.

Children's Test Anxiety Scale /CTAS

The “Children’s Test Anxiety Scale” developed by Wren and Benson (2004) is a 30-item measurement tool designed to assess test anxiety levels in children aged 8-12 in elementary and middle school. It is a four-point Likert-type scale (1 = Almost Never, 2 = Sometimes, 3 = Most of the Time, 4 = Almost Always) with a response time of 5-12 minutes. Based on the students' responses, the scale yields a score ranging from a minimum of 30 to a maximum of 120 points. Scores of 30–60 indicate low test anxiety, 60–90 indicate moderate test anxiety, and 90–120 indicate high test anxiety (Wren & Benson, 2004). The reliability estimate for CTAS is 0.92, and the Cronbach's Alpha values for the subscales Thoughts, Off-Duty Behaviors, and Autonomous Responses are calculated as .89, .78, and .85, respectively. The analyses showed that all item-factor loadings were statistically significant and ranged from 0.37 to 0.76 (more than half /0.60) (Wren & Benson, 2004). In the scale adaptation study conducted by Bulgan et al. (2017) in Türkiye, internal consistency was calculated using Cronbach's Alpha coefficient for reliability in terms of consistency. The Cronbach's Alpha value for the entire scale was .88, while the Cronbach's Alpha values for the subscales of Thoughts, Off-Duty Behaviors, and Autonomous Responses were .82, .70, and .75, respectively. Based on the analyses conducted, the validity evidence supports the construct validity of the 30-item, three-factor scale. The model demonstrated acceptable fit ($\chi^2/df = 3.97$, RMSEA = .05, SRMR = .05, CFI = .97, GFI = .92, AGFI = .90). Overall, the model fit indices indicate acceptable to good model fit, supporting the construct validity of the scale (Aydın & Bulgan, 2017). Below is a sample item administered in the classical format (Aydın & Bulgan, 2017), followed by its expanded version.

Classical form:

During exams...	Almost Never	Sometimes	Most of the Time	Almost Always
1. I wonder whether I will pass.	1	2	3	4

Expanded form:

Item 1:

- A. During exams, I almost never wonder whether I will pass.
- B. During exams, I sometimes wonder whether I will pass.
- C. During exams, I often wonder whether I will pass.
- D. During exams, I almost always wonder whether I will pass.

To collect qualitative data, 40-minute individual interviews using the think-aloud approach were conducted with 20 participants. Students were asked individually

to respond to statements using the classic Likert scale and to select response options on the extended-alternative Likert scale, and their responses were recorded. No other data collection tools were used for qualitative data.

Ethical Committee Approval

Letter dated July 11, 2023, No. 8543427050.04.04/990660, from the Chair of the Ethics Committee of the Rectorate of Ankara University.

Ministry of National Education, Directorate of Strategic Development, letter dated October 23, 2023, No. E-49614598-605.01-87840915.

Methodological Rigor

In the quantitative component of the study, Cronbach's alpha coefficients were calculated to assess the reliability of the measurement instruments, and confirmatory factor analysis (CFA) was conducted to obtain evidence of construct validity. In the qualitative component of the study, to enhance credibility, the coding performed by the first author during the data analysis process was reviewed by the second author, and the consistency of the coding was evaluated. In addition, to ensure that the findings accurately reflect the participants' views, direct quotes from the participants have been included.

Data Analysis

In the analysis of the data, respondents' answers were scored from 1 to 4 regardless of whether each item on the scale was positive or negative. The data obtained were analyzed using SPSS and Mplus software, employing methods appropriate to the research questions within the scope of the study. The assumptions of the tests were evaluated prior to the analyses. The significance level was set at .05 for the analyses. Before starting the analyses, descriptive statistics, missing data, and outliers were examined for data obtained from groups that received the classic Likert-type form, the extended-alternative form, or both, and the assumption of normality was checked. For the first and second research questions, the reliability of each form obtained from the extended-alternative and classic Likert-type scales was determined by calculating the Cronbach's Alpha internal consistency coefficient. Fisher's Z Test was used to examine whether the reliability coefficients obtained from the forms differed significantly from each other (Akhun, 1984). The comparison of different groups in this study is based on the assumption that errors are randomly distributed; this can be considered a limitation of the study. Validity evidence was obtained through CFA, and model fit values and factor loadings obtained from the extended-alternative and classic Likert-type scales were compared. Differences between item factor loadings were assessed using the independent-samples t-test. Within the scope of the third sub-objective, an independent-samples t-test was used to determine whether the total scores of respondents in the same group who completed both the classic and the extended-alternative forms of the entire scale differed from the total scores obtained from the scale's subscales, depending on whether they completed the

extended-alternative Likert-type scales. The profiles (distribution of responses to items) of different individuals who obtained the same total score on the expanded-alternative Likert-type scale and those of individuals who obtained the same total score on the classic Likert-type scale were examined descriptively at the item and subscale levels, and the differences were presented. To determine how students interpreted the statements in the classic form and the different response options in the extended-alternative form (within the scope of the fourth sub-objective), interviews with the students were transcribed and subjected to content analysis. The results were presented using themes, categories, and codes, accompanied by quotations illustrating students' views.

Results

The primary objective of this study is to examine evidence of validity and reliability obtained from extended-alternative Likert-type scales and classic Likert-type scales applied to different groups, and to determine whether there are differences in reliability coefficients, fit indices, and item factor loadings. To determine the reliability of the extended-alternative and classic Likert-type scales applied to different groups, Cronbach's Alpha internal consistency coefficients were calculated. Subsequently, Fisher's Z Test was employed to test the differences between these reliability coefficients. The Cronbach's alpha reliability coefficients for the extended-alternative and classic Likert-type scale forms are provided in Appendix 1. The results of Fisher's Z test assessing whether the obtained reliability coefficients differed significantly from each other are presented in Table 1.

Table 1

Comparison of Form Reliability Results

	Classic Form	Extended-Alternative Form	Zrclassic	Zrextended	Z
All Scale	.88	.87	1.37	1.35	.30
Thoughts	.83	.81	1.18	1.14	.64
Off-Duty Conduct	.64	.61	.77	.74	.42
Autonomous	.76	.78	1.00	1.05	-.70

As shown in Table 1, since all z values are less than 1.96 (Akhun, 1984), there is no significant difference between the reliability coefficients of the classic and extended-alternative Likert-type scales. In other words, converting the classic Likert-type scale to an extended-alternative form did not create a difference in either the reliability coefficient obtained from the entire scale or the reliability coefficients obtained from the scale's sub-dimensions. Validity evidence was obtained through confirmatory factor analysis, and reliability coefficients for the extended-alternative and classic Likert-type scales were compared. For evidence of validity, confirmatory

factor analysis was performed, and model-fit indices and factor loadings from the expanded- alternative and classic Likert-type scales were compared. The data fit indices and factor loadings obtained from the classic Likert-type scale form and the expanded alternative Likert-type scale form were compared. Data fit values and factor loadings obtained from the classical Likert-type scale and the extended-alternative Likert-type scale form were examined for goodness of fit using confirmatory factor analysis and are shown in Table 2.

Table 2

Model Fit for the Classic Likert-Type and Extended-Alternative Likert-Type Scale Form

Test Form	χ^2	χ^2/sd	RMSEA	SRMR	CFI
ClassicForm	5507.05	12.65	.05	.06	.89
Extended- Alternative Form	5627.08	12.9	.04	.05	.93

For fit indices, the ideal value for CFI should be greater than .90, while the RMSEA value should be less than .08 (Uyumaz & Çokluk, 2016). When examining the goodness-of-fit values obtained from the scales, an χ^2/sd value not below 5 indicates that the model fit is not good. The RMSEA value for the classic form is .05, indicating a good fit, while the RMSEA for the extended-alternative form is .04, indicating an excellent fit. In terms of SRMR values, the classic form has a value of .06 and the extended-alternative form has a value of .05, both of which indicate a good fit. The CFI value obtained from the classic form is .89, indicating that it is very close to a good fit. When the fit indices CFI, SRMR, and RMSEA are considered together, confirmatory factor analysis indicates that the three-factor scale structure is acceptable and valid. In their study, Zhang and Savelei (2016) considered the chi-square, CFI, and RMSEA values when comparing the model fits of the classic and extended-alternative forms and found that the extended-alternative form had higher fit values. In this study, consistent with the relevant finding, it was determined that there were minor differences between the forms and that the extended-alternative form had a better fit. Thus, it is consistent with the findings of Zhang and Savalei. Examination path diagrams related to confirmatory factor analysis obtained from the group to which the original form was applied revealed that all items belonging to the expanded-alternative form had factor loadings greater than .30. In the classic form, all items except item 18 had factor loadings greater than .30 and showed good fit with the three-factor structure. Kam et al. (2024) examined the factor loadings of the scale items for the extended form using CFA; in the present study, the same factor loadings were also examined using CFA. The factor loadings of all items were quite high and were loaded onto the same factor as in the original structure. The factor loadings of the items in the two different scales were compared using a paired- samples t-test, and the results are presented in Table 3.

Table 3

Comparison of Factor Loadings of Items on the Scale Across Forms Using an Paired Samples t-Test (N: 30)

Forms			M	Sd	T	p
Classic-Extended-Alternative	Form	Factor	.00	.71	.43	.66
Loads						

As shown in Table 3, when comparing the factor loadings of the classic Likert-type and extended-alternative Likert-type scale forms, the t-test result indicates a significance level ($p = .66$) that is greater than .05 ($t(29) = 0.43$, $p = .66$). Therefore, it can be stated that the factor loadings of the items in the classic and extended-alternative forms of the scale do not differ significantly. The analysis concluded that the model fits for the extended-alternative forms showed better fit than those obtained for the classical forms. Furthermore, when comparing the item factor loadings grouped under the factors, it was found that the item loadings in the extended-alternative form were higher than those in the classical form, indicating better construct validity.

Within the scope of the second sub-objective of the study, the validity and reliability evidence obtained from extended-alternative and classic Likert-type scales administered to different groups was examined separately to determine whether the evidence differed between the child and adolescent groups. The reliability of the scores obtained for the extended-alternative Likert-type form was examined using the Cronbach Alpha internal consistency coefficient. The Cronbach's alpha for the adolescent group ($\alpha = .91$) was higher than for the child group ($\alpha = .86$). The Cronbach's alpha values for the classic form obtained from the adolescent group ($\alpha = .84$) and the child group ($\alpha = .89$) were examined. In this case, the Cronbach's Alpha value for the adolescent group was lower than that for the child group. The results are shown in Appendix 2. Fisher's Z test was used to determine whether the reliability coefficients for the child and adolescent groups differed significantly. The results of Fisher's Z test for the adolescent and child groups are presented in Table 4.

Table 4

Comparison of Reliability Results for Children and Adolescent Respondents in Different Groups

	Classic Form					Extended-Alternative Form				
	C	A	Zrc	Zra	Z	C	A	Zrc	Zra	Z
All Scale	.89	.84	1.42	1.25	1.63	.86	.91	1.29	1.52	-2.26*
Thoughts	.82	.83	1.17	1.19	-.15	.78	.87	1.05	1.33	-2.64*
Off-Duty Conduct	.68	.55	.83	.62	2.01*	.61	.70	.71	.87	-1.54
Autonomous Responses	.76	.75	1.01	.98	.33	.75	.84	.98	.88	.95

*p<.05. Note: (C=Child Group; A=Adolescent Group)

As shown in Table 4, the z-values of respondents who answered only the extended-alternative Likert-type scale form are smaller than the test value of 1.96 for the off-duty and autonomous responses sub-dimensions; therefore, there is no significant difference between the reliability coefficients of the scales in the child and adolescent groups. When comparing reliability values for all scales and for the thoughts dimension, the reliability coefficients differ significantly between the child and adolescent groups. Examination of the Fisher z-values for respondents who answered only the classic Likert-type scale revealed no significant difference between the reliability coefficients of the scales for the child and adolescent groups for the thoughts and autonomous reactions sub dimensions. However, comparing the reliability values for all scales and for the task-irrelevant reactions dimension indicates a significant difference between the scales' reliability coefficients in the child and adolescent groups. For validity evidence, Table 5 shows the model-fit values from the confirmatory factor analysis conducted on data from the extended-alternative form for the child and adolescent groups.

Table 5

Expanded-Alternative Form Fit Indices for Child and Adolescent Respondents in Different Groups

Compatibility Indexes		χ^2	χ^2/df	RMSEA	SRMR	CFI
Extended-Alternative Form	Child Group	574.08	1.42	.04	.06	.93
	Adolescent Group	549.04	1.36	.04	.07	.93
Classic Form	Child Group	758.48	1.88	.04	.07	.90
	Adolescent Group	531.92	1.32	.03	.09	.87

When the goodness-of-fit values for the extended-alternative Likert-type form are examined, the RMSEA indicates excellent fit for both the child group (RMSEA=.04) and the adolescent group (RMSEA=.04). Examination of the SRMR values for the scale forms, shows that both the child and adolescent groups exhibit good fit. When examining of the CFI values for the extended-alternative scale form, it is seen that they correspond to a good fit in both the child and adolescent groups (CFI=.93). When overall fit indices are evaluated, confirmatory factor analysis of data from child and adolescent groups indicates that the three-factor scale structure yields acceptable and valid results. When examining the goodness-of-fit values obtained from the children and adolescent groups for the classic Likert-type scale form, RMSEA indicates excellent fit for both the children (RMSEA=.04) and adolescent (RMSEA=.03) groups. The SRMR values for the classic Likert-type scale form showed a good fit in the child group (SRMR=.07) and in the adolescent group (SRMR=.09). The CFI indicates a good fit in the child group (CFI=.90) and in the adolescent group (CFI=.87). When the fit indices are evaluated overall, the confirmatory factor analysis of data obtained from child and adolescent groups indicates that the three-factor scale structure yields acceptable and valid outcomes. With respect to classical form fit values, the child group exhibited better fit values than the adolescent group. Since some fit values indicated better fit in the child group and others indicated better fit in the adolescent group, a clear comparison could not be made. The factor loadings of the items in the scales were compared using the independent samples t-test, and the results are presented in Table 6.

Table 6

Comparison of Factor Loadings for Items Included in Responses from Children and Adolescents in Different Groups (N: 30)

	M	Sd	T	p
Child-Adolescent Factor Loadings for the Extended-Alternative Form	-.06	.08	-3.94	.001
Child-Adolescent Factor Loadings for the Classic Form	-.03	.11	-1.48	.14

As shown in Table 6, comparison of the factor loadings of the expanded-alternative Likert-type scale for the child and adolescent groups indicates a significant difference between the item factor loadings for the child and adolescent groups ($t(29)=-3.94$, $p<.05$). The effect size (Cohen's $d = 0.72$) indicates that this difference is of moderate-to-high magnitude and that the age group has a significant effect on factor loadings. When comparing the factor loadings of the classic Likert-type scale between the child and adolescent groups, the significance level associated with the t value ($p=.14$) is greater than .05, indicating that there is no significant difference between the factor loadings of the items in the scale for the child group and the adolescent group ($t(29)=-1.48$, $p>.05$). Examination of the factor loadings of items obtained from

the children and adolescent groups shows that approximately half of the items in the thoughts sub-dimension have higher loadings in the children group than in the adolescent group. The average factor loadings for the forms in the child and adolescent groups are shown in Table 7 for the scale and its sub-dimensions.

Table 7

Average Factor Loadings for Forms from Child and Adolescent Respondents in Different Groups

	Child Group		Adolescent Group	
	Classic Form	Extended-Alternative Form	Classic Form	Extended-Alternative Form
All Scale	.56	.52	.53	.58
Thoughts	.57	.54	.57	.61
Off-Duty Conduct	.53	.47	.44	.49
Autonomous Responses	.58	.55	.55	.64

As shown in Table 7, the factor loadings of the classic form are higher in the child group, whereas the average factor loadings of the expanded-alternative form are higher in the adolescent group, both based on scales and subscales. In this case, expanding the scale can be interpreted as useful for adolescents. When factor loadings were examined for each item across both forms, it was found that Item 1 and Item 29 in the “thoughts” subscale, as well as Item 7 and Item 18 in the “off-task behaviors” subscale, had low loadings in the child group across both forms.

Within the scope of the third sub-objective of the study, an independent samples t-test was used to analyze whether the total scores of respondents within the same group differed according to their status of taking extended-alternative Likert-type scales; the analysis results are presented in Table 8.

Table 8

Comparison of Scale Total Scores According to Different Scaling Conditions

Scale Form	N	M	sd	<i>t</i>	<i>p</i>
Extended- Alternative Form	97	68.38	16.14	-2.86	.00
Classic Form	97	70.55	16.84		

As shown in Table 8, a related-samples t-test was used to the total scores of individuals in the group to which both the classic and extended-alternative Likert-type scale forms were applied. As a result of the analysis, the *t* value was statistically significant ($t(96) = -2.87, p < .05$). Therefore, the students' total scale scores differed,

favoring the extended-alternative form, depending on the type of scale used. The effect size ($d = 0.29$) indicates that the difference is small and the scale format has a limited effect on total scores.

For the study's fourth sub-objective, respondents' answers to the items were presented using themes, categories, and codes, distinguishing between responses consistent with the test's purpose and those inconsistent with it, to examine how they interpreted the different response options. Students' responses to the items in the classic form were examined as responses that were either in line with or not in line with the purpose of the test, and their comments on the items were organized into five categories and 13 codes under three themes. The quotes from the respondents are presented in Appendix 3. When the personal characteristics category in Appendix 3 was examined, students gave responses that were not appropriate for the purpose of the code for cheating when reviewing their answers to the statements 'I look at my friends during exams' and 'I look at what is going on around me during exams' on the scale. Regarding time management, the responding students provided answers appropriate for the test's purpose. It was observed that the students explained the statements "I am afraid of doing something wrong during the exam," "I think about other things during the exam," and "I get distracted during the exam" using the attention code and did not provide answers appropriate to the test's purpose.

The first code in the emotional characteristics category of the personal characteristics theme is "test anxiety/fear," and when attention is paid to the students' responses, participants provided answers appropriate to the purpose of the test. The second code in this category is "excitement," and it was observed that the students' responses were appropriate for the purpose of the test. The third code in this category is "boredom," and looking at the students' responses, it can be concluded that the test served a purpose other than its intended one. Within the theme of exam characteristics, it was observed that students interpreted the concept of exam difficulty differently, shaped their perception of exam difficulty according to their own success, and associated the exam with a single course, answering only with respect to that course. In subsequent studies, the items should be standardized by eliminating their ambiguities. Regarding family characteristics, students' anxiety about the exam may be related to their family's attitude. Based on students' responses, this theme was clear and straightforward: students understood it easily and did not assign different meanings to it. Students' responses to the items in the extended form were examined as appropriate or not appropriate for the purpose of the test, and their comments on the items were categorized into five categories under three themes and 13 codes, which are presented in Appendix 4.

The theme of personal characteristics consists of three categories and 10 codes: personal, cognitive, and affective characteristics. First, the personal characteristics category was examined, revealing that these students provided responses inappropriate for the intended purpose. Students' statements about playing with their pencils during the exam, having headaches, or shaking their feet, which they thought

about to remember the answers, were grouped under the code for retrieving information. When the findings grouped under the theme of personal characteristics were examined, students answered most of the items appropriately; student responses were similar to each other, response diversity was lower than in the classic form, and the answers for the items were predominantly clustered in certain categories. In this case, it was observed that the items in the expanded-alternative form were more understandable and clearer than those in the classic form, and even when students interpreted the items differently, the differences were minimal. A category related to exam difficulty has been created with two codes: 'the difficulty of the exam' and 'the ease of the exam'. In these codes, under the ease of the exam, the response given by the second student to the statement "I try to finish the exam quickly" is quoted. Under the exam difficulty code, the responses given by student number one ("I have trouble remembering the answers"), student number seven ("I wonder if my answers are correct"), and student number ten ("I'm afraid of failing the exam") are included in the excerpt section. When the responses of students who completed the extended form were considered, it was observed that they provided answers to some statements inconsistent with the scale's purpose and that the test served a purpose other than its intended one. When considering the students' statements, it was observed that they generally marked "always" or "never" for statements such as "I play with my pen, I have a strange pain in my stomach, I check the time, I find it difficult to sit still, I think I should have studied more, I feel the need to go to the bathroom". In their responses to statements such as "I look around at what is happening around me," "I feel anxious," "I wonder if my answers are correct," and "I try to finish my exam quickly," it was observed that they responded based on the action element. In the responses received from students regarding the items on the scale, it is noteworthy that all students responded to the item "I feel the need to go to the bathroom during exams" in relation to time, while all students responded to the item "I get lost in thought during exams" in relation to action.

Discussion, Conclusion and Suggestions

This study compared the psychometric properties resulting from converting classic Likert-type scales to extended alternative Likert-type scales. The study examined the effect of the extended alternative Likert-type scale form on validity and reliability across different age groups (children and adolescents), and whether the total scale score differed between the extended alternative and classic Likert-type scales. In this context, the test anxiety scale for children—initially in the classic Likert-type scale format—was converted to an expanded format based on responses obtained from children, expert opinions, and the think-aloud method, and administered to students in grades 3 through 6. The results obtained from this study are presented below.

In line with the first sub-objective of the study, it was observed that reliability evidence obtained from the extended-alternative Likert-type scales from different groups and the reliability evidence obtained from the classic Likert-type scales were

high in both forms, but there was no statistically significant difference between the two forms. Zhang et al. (2023) applied both the classic and extended-alternative Likert-type scale forms of the Rosenberg Self-Esteem Scale to an adult group in their study, and the Cronbach's Alpha value for the classic form was calculated to be close to that of the extended-alternative form. Converting the classic Likert-type form to the extended-alternative Likert-type form did not affect the reliability coefficient for the entire scale or the reliability coefficients for the scale's sub dimensions. When both forms were considered, the values obtained for the off-duty behaviors sub-dimension were lower than those for the original form. This situation is thought to stem from items in the sub-dimension that were not clear and concise, limited reading comprehension skills among the children, and the inability of students from different cultural backgrounds and younger age groups in the study group to fully understand the concepts in the statements. For example, it was observed that younger individuals and students from different cultures did not fully understand the expression "my face doesn't get hot" in the statement "my face doesn't get hot."

When examining the validity evidence obtained from extended-alternative Likert-type scales and the validity evidence obtained from classical Likert-type scales, the fit indices and factor loadings were compared, and it was concluded that all items in both forms were meaningful and that the extended-alternative form could be a model with better model fit values. In this study, consistent with the relevant finding, it was determined that minor differences existed between the forms and that the extended-alternative form provided a better fit; this result was consistent with the findings of Zhang and Savalei (2016).

Examination of the validity evidence from extended alternative Likert-type scales and from classical Likert-type scales, based on CFA results, indicated that both scale forms possess construct validity owing to high factor loadings of items in the scales and their subscales; however, no significant difference was found between the two forms. Zhang and Savalei (2016) used the Beck Depression Scale and the Rosenberg Self-Esteem Scale in their study and found that the factor structure of the scales in the extended format was better than that of the scales in the classic Likert format using CFA. In this study, the Test Anxiety Scale for Children was used. The findings suggest that the two scales have a similar factor structure and that this similarity may stem from the measurement tool used and from the characteristics of the participants to whom the measurement was applied. When examining average factor loadings and item factor loadings in both the classical and extended-alternative forms separately, it was observed that results varied by sub dimension.

In line with the second sub-objective of the study, the reliability evidence from the classic Likert-type and extended-alternative Likert-type scale forms were examined across child and adolescent groups. Considering the z-values obtained from the classic form for the child and adolescent groups, there was a significant difference between the reliability coefficients in the task-related behaviors sub-dimension. For the extended-alternative form, the z-values obtained from the child and adolescent

groups showed a significant difference in reliability coefficients for the entire scale and the thoughts sub dimension. However, no significant difference was observed between the child and adolescent groups in the autonomous reactions sub dimension for either form. In the off-task responses subscale, there was a significant difference between the reliability coefficients of the classic form for the child and adolescent groups, whereas no significant difference was observed for the expanded-alternative form. This situation may be attributable to the clarification of the items and the consequent reduction in confusion. On the thoughts subscale, the child and adolescent groups differed significantly on the extended-alternative form but not on the classic form. On the thoughts subscale, the adolescent group responded to the items presented in the extended-alternative format in a manner more consistent with the scale's intended meaning.

It is believed that this situation stems from the fact that the child group is in the concrete operational stage and that they have difficulty understanding abstract expressions. Finally, examination of the autonomy responses subscale revealed no significant difference between the reliability coefficients of the child and adolescent groups for both the classical and extended-alternative forms. In this study, the classical form was relatively more reliable for the child group, whereas the extended-alternative form was more reliable for the adolescent group. This difference may be attributed to the study group and the data-collection tool, because the longer items in the extended-alternative form increased the reading load, thereby resulting in more random measurement errors affecting the child groups. When the validity evidence obtained from the classical and extended-alternative Likert-type scales from different groups was examined using fit indices for the child and adolescent groups, the extended-alternative form had better fit indices for the child group, whereas the classical form had better fit indices for the adolescent group. Examination of the factor loadings for the forms administered to the groups showed that all items had statistically significant loadings; given these high loadings and the model fit indices, both the classical and the extended-alternative demonstrate structural validity for scores obtained from the child and adolescent groups. Comparison of the factor loadings of the classical and extended-alternative Likert-type scales between the child and adolescent groups revealed a significant difference favoring the adolescent group for the extended-alternative form, whereas no significant difference was observed for the classical form.

A significant difference was found between the total scores obtained from the classical and extended-alternative Likert-type scales for the same individuals within the same group, with the total scores from the classical form being lower than those from the extended-alternative form. This result is likely due to expanding the scale items and specifying their frequency more clearly. Additionally, the small time difference between the application of the classical and extended-alternative forms is thought to result from individuals remembering their responses to the items rather than from the short interval between two applications.

Although the students had the same scale score, they gave different responses to the items. Therefore, individual interviews were more effective than using different forms of the scale for distinguishing individual profiles and for assessing whether the scale items serve their intended purpose. In the study by Zhang et al. (2019), the vagueness and ambiguity of the items in the classical form were highlighted as disadvantages. Examination of themes derived from student responses in this study revealed that some items in the classical form were not sufficiently clear or precise, leading students to provide different responses to those items. This situation pertains more to the clarity and comprehensibility of the items than to the scaling process and indicates that all of them should be evaluated together.

Taken all together, the findings indicate that the classic Likert-type scale and the expanded-alternative Likert-type scale form exhibit similar psychometric properties in terms of reliability and general construct validity. This suggests that the classic Likert-type form can continue to be used for general measurement purposes. However, when considering age-group fit indices, the extended-alternative form's better fit in the child group suggests that expressing items more clearly and concretely could improve measurement quality in this age group. Conversely, the fact that the classic form shows a better fit in the adolescent group suggests that this group is capable of interpreting more abstract and general statements. Significant differences in total scores obtained from the same individuals, coupled with lower scores on the classic form, indicate that the item format influences individuals' response behavior. This finding also indicates that scores obtained from scales with different formats should not be directly compared. The fact that individuals' responses to items differ, even when their total scores are similar, suggests that scale formats alone may not be sufficient, particularly when detailed individual-level evaluations are desired. Therefore, when the goal is to gain a deeper understanding of an individual's profile, scale-based assessments should be supplemented with individual interviews.

However, the findings clearly highlight the importance of systematically examining how scale items are understood by the target age group during scale development and adaptation studies. It is recommended that pilot studies—especially those using the think-aloud method—be conducted with a small sample before administering the scale items to the target age group. This study shows that the meanings students visualize when reading relevant statements can vary. This situation demonstrates that linguistic clarity alone is insufficient; cognitive processes must also be taken into account. Consequently, it is believed that qualitative pilot studies that reveal how scale items are interpreted by the target audience constitute a critical step in strengthening the validity evidence of the measurement instrument.

Based on the findings of this study, it is recommended that the explanatory texts in scale applications be shortened and clarified. Conducting psychometric comparisons of different rating scales (3-point, 5-point, 7-point, etc.), numbers of items (short and long forms), and response categories is important for future studies. In light of the findings, researchers using the classic form should exercise caution

because some items may be interpreted differently, whereas the extended-alternative form is considered appropriate for groups with well-developed abstract thinking skills (adolescents). Additionally, qualitative approaches, such as individual interviews and the think-aloud method, make significant contributions to evaluating how items are understood and to assessing construct validity. Finally, due to the differences in factor loadings between child and adolescent groups, it is recommended that measurement invariance be tested in future studies.



Psikolojik Testlerde Genişletilmiş Likert Tipi Ölçek ile Klasik Likert Tipi Ölçeklerin Psikometrik Özelliklerinin Karşılaştırılması*

•Seda Gözlüöl¹, •Seher Yalçın²

1 Aile ve Sosyal Hizmetler Bakanlığı, Gaziantep, Türkiye

2 Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi Müdürü, Ankara Üniversitesi, Ankara, Türkiye

ÖZ

Bu çalışmada klasik Likert tipi ölçekle alan yazında yeni geliştirilen genişletilmiş alternatif Likert tipi ölçeklerin psikometrik özelliklerinin; farklı ölçek alma durumlarına bağlı toplam puanlarda ve genişletilmiş-alternatif formun çocuk ve ergen grupları için psikometrik özelliklerinin karşılaştırılması amaçlanmıştır. Karma modellerden açıklayıcı desende tasarlanan çalışmada, nicel veriler toplanıp analiz edilmiş daha sonra nicel verileri açıklamak ve ayrıntılı olarak incelemek için nitel veriler toplanmış ve analiz edilmiştir. Çalışmada amaçlı örnekleme ve kolay erişilebilir örnekleme yöntemleri kullanılmıştır. Çalışma grubunu Gaziantep ve Kilis ilinden 2023-2024 eğitim-öğretim yılı ilkökul ve ortaokul öğrencileri oluşturmaktadır. Nicel verilerin toplanmasında 892 öğrenci ve nitel görüşmeler için 20 öğrenci çalışmaya katılmıştır. Araştırmada ‘Çocuklar İçin Sınav Kaygısı Ölçeği (ÇSKÖ) kullanılmıştır. Formlara ait geçerlilik ve güvenirlik kanıtları arasında manidar fark olmadığı görülmüştür. Cevaplayıcıların toplam puanlarının aldığı ölçek formuna göre farklılaşıp farklılaşmadığı incelenmiş olup aynı gruptan alınan toplam puanların aldığı ölçek formuna göre formlar arasındaki farkın manidar olduğu görülmüştür. Genişletilmiş-alternatif formun, öğrencilerin maddeleri daha tutarlı ve amaca uygun yorumlamasını sağladığı, ancak her iki formda da bazı maddelerin farklı bilişsel süreçler üzerinden yanıtlandığı belirlenmiştir.

Anahtar sözcükler: likert tipi ölçek, genişletilmiş-alternatif likert tipi ölçek, güvenirlik kestirimleri, geçerlilik kanıtları, sesli düşünme.

*Bu makale, Seda Gözlüöl’ün, Prof. Dr. Seher Yalçın Danışmanlığında Ankara Üniversitesi Eğitim Bilimleri Enstitüsünde 2024 yılında tamamladığı “Psikolojik Testlerde Genişletilmiş Likert Tipi Ölçek İle Klasik Likert Tipi Ölçeklerin Psikometrik Özelliklerinin Karşılaştırılması” başlıklı Yüksek Lisans tezinden üretilmiştir.

1 Sorumlu Yazar: Öğretmen, E-posta: sedaakrldg@gmail.com ORCID ID: <https://orcid.org/0009-0003-1405-7550> ROR ID: <https://ror.org/04n9e8n41>

2 Prof. Dr., E-posta: yalcins@ankara.edu.tr, ORCID ID: <https://orcid.org/0000-0003-0177-6727>, ROR ID: <https://ror.org/01wntqw50>

Gönderim Tarihi: 03.11.2025 Kabul Tarihi: 02.06.2026 Yayın Tarihi: 06.08.2026

Atıf: Gözlüöl, S., & Yalçın, S. (2026). Psikolojik testlerde genişletilmiş likert tipi ölçek ile klasik likert tipi ölçeklerin psikometrik özelliklerinin karşılaştırılması. *Ankara University Journal of Faculty of Educational Sciences*, 59(2), 591-638. <https://doi.org/10.30964/auebfd.1816495>

Makaleye ilişkin tüm etik beyanlar, beyanlar sayfasında yer almaktadır (Sayfa: 632).

Psikolojik testler; insanların davranışlarını, performanslarını, tutumlarını, yeteneklerini kısaca tüm psikolojik özelliklerini ölçmek ve var olan durumu ortaya koymak amacıyla oluşturulmuş uzmanlığı almış kişiler tarafından standart biçimde kullanılabilen ölçme aracıdır (Cansız, 2007; Kutanis & Elçi, 2006). Kişilere ilişkin doğru karar vermek; tanılama, seçme ve yerleştirme yapmak, aynı zamanda bireysel farklılıkları ya da aynı kişinin farklı zamanlardaki değişim ve gelişimini ortaya koymak amacıyla da kullanıldığında testin uygulandığı bireyin durumunu doğru bir şekilde ortaya koyması, kullanılan ölçme aracının amaca uygun olmasına, ölçme yapılan ortama, ölçmeyi yapan kişinin uzmanlığına ve ölçme araçlarının güvenilirlik, geçerlilik, kullanılabilirlik gibi psikometrik özelliklerine bağlıdır (Börkan ve diğ., 2017; Cohen & Swerdlik, 2018; Şahin & Gülleroğlu, 2013; Tülübaş, 2009).

Klasik Likert tipi ölçeklerde, cevaplayıcının ifadeleri görüş, tutum ve duygularını tek tek yansıttığı için ifadelerden elde edilen puanların toplanarak bireylerin verdikleri cevaplardan bir toplam puan belirlendiği, dereceleme toplamalarıyla ölçekleme yapılı (Gökdemir & Yılmaz, 2023; Şahin & Gülleroğlu, 2013; Turan ve diğ., 2015). Klasik Likert tipi ölçeklerin psikometrik özelliklerinin yüksek olması, ölçekten elde edilen toplam puanların, ölçeğin alt boyutlarından elde edilen toplam puanlarla karşılaştırılabilir olması, standardize ölçümlerle veri toplanması ve tekrarlı ölçümler yapılarak geçerlilik-güvenirlik analizlerinin yapılabilmesi gibi avantajlara sahip olması açısından birçok alanda kullanılmaktadır (Gökdemir & Yılmaz, 2023). Ancak klasik Likert tipi ölçeklerin avantajlarının yanında sınırlılıkları ve dezavantajları da bulunmaktadır. Son zamanlarda klasik Likert tipi ölçeklerin, çocukların değerlendirilmesinde, tanılanmasında ve tedavisinde de sıklıkla (Roeen, 1985; Uyumaz & Çokluk, 2016) kullanılmasına rağmen klasik Likert tipi ölçeklerin psikometrik özelliklerinin analiz edilmesinde çoğunlukla yetişkin grubunun değerlendirmeye alındığı, bu durum klasik Likert tipi ölçeklerin küçük yaş gruplarında kullanımını, psikometrik özelliklerini şüpheye düşürmektedir (Alan, 2019; Borgers & Hox, 2001).

Klasik Likert tipi ölçekte yer alan maddeler ve yanıt seçenekleri üzerindeki anlaşmanın derece belirsizliği, maddelerin herkes tarafından aynı şekilde anlaşılabilmesi, ölçekte yer alan maddelerin farklı bireyler tarafından farklı yorumlamalara neden olması, maddenin neyi kastettiğinin (Sıklık mı? Derece mi?) muğlak olması, cevaplayıcının görüş ve tutumunu tam yansıtamaması ve yanıt kategorisindeki derecelemelerin karşılık geldiği miktarın kişiden kişiye farklılaşması Likert tipi ölçeğin sınırlı yanlarından olup psikometrik özelliklerinin düşmesine neden olmaktadır (Başaran, 1997; Joshi vd., 2015; Kam vd., 2024). Klasik Likert tipi ölçeklerin dezavantajlarından biri de farklı cevap ifadeleri ile aynı toplam puanı alan bireylerin aldıkları toplam puanların birbirleriyle aynı olması durumu olup aynı toplam puanı alan bireylerin aynı profillere sahip olunduğu düşüncesidir. Bireylerin birbirinden ne kadar farklılaştığını da ortaya koyamaması klasik Likert tipi ölçeğin dezavantajıdır (Alan, 2019). Klasik Likert tipi ölçmelerin geçerlik ve güvenilirliğine ilişkin son yıllarda yapılan çalışmalardan birinde (Zhang & Savalei, 2016), klasik Likert tipi ölçeklere yeni bir boyut kazandırmışlardır. Araştırmacılar, klasik Likert tipi

ölçeklerde yer alan maddelerin katılıyorum-bazen-katılmıyorum şeklinde yanıtlanmasının davranışın sıklığının ya da derecesinin muğlak bir ifade olduğunu bu nedenle cevaplayıcı için çok özel sonuçlar vermediğini belirtmiştir. Araştırmacılar, her yanıt seçeneği yerine tam bir cümle kullanıldığı ve maddedeki muğlak ifadelerin giderilmeye, azaltılmaya çalışıldığı, genişletilmiş biçim adı verilen genişletilmiş-alternatif bir ölçek biçimi önermişlerdir.

Genişletilmiş-alternatif Likert tipi ölçekler ilk olarak 2002 yılında Hills ve Arglye tarafından öne sürülmüş olup “The Oxford Happiness Questionnaire: A compact scale for the measurement of psychological well-being” adlı çalışmasında Oxford mutluluk ölçeği üzerinden yola çıkarak cevaplayıcının ölçekte yer alan yanıt seçeneklerinde katılıyorum’dan katılmıyorum’a doğru sıralanan dereceleme ifadelerini işaretlemesinden ziyade her dereceleme ifadesini bir cümle halinde sunulması esasına dayanarak genişletilmiş Likert tipi ölçeği geliştirilmişlerdir (Hills & Arglye, 2002).

Zhang ve Savalei (2016) yaptığı çalışmada, genişletilmiş-alternatif Likert tipi ölçeklerin psikometrik özelliklerinin daha yüksek olduğunu ve bireylere ait profilleri daha iyi açıkladığını öne sürmüştür. Genişletilmiş-alternatif form, cevaplayıcının maddeleri daha dikkatli okuması, seçenekler arasındaki farkı kolayca anlamasını sağlayarak cevaplayıcının kafa karışıklığını azaltır ve daha dikkatli cevaplama yardımcı olurken klasik Likert formdaki ölçeklerde yer alan maddelere, cevaplayıcının ne düzeyde katılıp katılmadığı belirsiz olup değerlendirmeye açık değildir (Kam ve diğ., 2024). Klasik Likert tipi ölçekte cevaplayıcı katılmıyorum/katılıyorum seçeneklerinden birini işaretlemek zorundayken genişletilmiş-alternatif form maddelerin cevaplayıcıların herhangi bir ifadeye katılmasını gerektirmez, katılmıyorum-katılıyorum seçeneklerinden kurtarır ve rıza yanlılığı teorik olarak ortadan kaldırılır. Cevaplayıcılar, ölçekte yer alan maddeleri cevaplarken olumlu maddelerden olumsuz maddelere geçiş olmaması ve her maddeye ilişkin yanıt seçeneklerinin cümle ile belirtilmesiyle herhangi bir karışıklık yaşamayacaklardır. Cevaplayıcıların önyargısını azaltır. Klasik Likert tipi ölçekte maddelerin olumlu/olumsuz madde yönü varken genişletilmiş-alternatif formdaki maddelerin yönü yoktur, her maddenin hem olumlu hem de olumsuz versiyonları yanıt seçenekleri olarak sunulmaktadır (Zhang ve diğ., 2019; Zhang ve diğ., 2023).

Genişletilmiş-alternatif Likert tipi ölçekte yer alan maddelerin cevaplanması, klasik Likert tipi ölçek maddelerine göre daha fazla okuma yükü gerektirir. Genişletilmiş- alternatif Likert tipi ölçekte maddelerin uzunluğundan kaynaklı cevaplayıcılarda yorgunluk, sıkılğanlık ve dikkatsizlik nedeniyle istenmeyen yanıtlamalar meydana gelebilir bu nedenle cevaplayıcının maddeleri okurken daha dikkatli olması gerekmektedir. Bu durum ölçeğin psikometrik özelliklerini düşürebilir. Genişletilmiş- alternatif Likert tipi ölçeğin geliştirilmesi, gruba uygulanması ve değerlendirilmesinin; cevaplayıcıların ölçekte yer alan maddeleri doldurulmasının klasik Likert tipi formdan daha uzun sürmesi nedeniyle zaman ve maliyet açısından; ekonomik değildir (Zhang ve diğ., 2019). Klasik Likert tipi ölçek

maddelerinin, genişletilmiş-alternatif Likert tipi formunda maddelere dönüştürülmesiyle hem araştırmacılar hem de cevaplayıcılar, maddelerin yanıt seçeneklerinin amaçlanan anlamını daha net anlayacakları, ifadelerin bireyler tarafından farklı yorumlanma olasılığını azaltacağı ve ölçeğin ölçmek istediği boyutu daha geçerli bir şekilde ortaya koyacağı düşünülmektedir. Bu nedenle klasik Likert formdaki ölçeklerin yerine alternatif formların kullanılması Zhang ve Savalei (2016) tarafından önerilmiştir.

Alan yazında genişletilmiş-alternatif Likert tipi ölçeklerle ilgili araştırmaların bahsi geçen araştırmacılarla ve yetişkin grubu üzerinden yapılan çalışmalarla sınırlı kaldığı ve Türkiye’de ise genişletilmiş-alternatif Likert tipi ölçeğe ait bir çalışmanın bulunmadığı görülmektedir. Bu nedenle klasik ve genişletilmiş-alternatif Likert tipi ölçeklerin psikometrik özelliklerinin incelenmesi ve karşılaştırılması, alternatif Likert tipi ölçeklerde; kişilere ait profillerin daha açık bir şekilde ortaya konulması açısından önemli olduğu düşünülmektedir. Bu çalışmada Likert tipi bir psikolojik testin genişletilmiş-alternatif Likert tipi ölçeklere çevrilerek psikometrik özellikler bakımından karşılaştırılması ve çocuk grubunda ölçek anlaşılabilirliği ile ilgi sorunların daha fazla olması nedeniyle geçerlik güvenirlik farklılıklarının çocuk ve ergen grubuna göre de ayrı ayrı incelenmesine gereklilik görülmüştür. Orijinalinde klasik Likert tipi olarak derecelendirilen çocuklar için sınav kaygısı ölçeğinin genişletilmiş-alternatif Likert tipi ölçeklere çevrilerek psikometrik özellikler bakımından karşılaştırılması, genişletilmiş formda farklı yaş gruplarında (çocuk-ergen grup) olmanın geçerlik ve güvenirliğe etkisinin incelenmesi amaçlanmıştır. Ayrıca ölçek toplam puanı ve öğrenci profillerinin (ölçek alt boyutlarındaki ölçülen özelliğe düşük, orta ve yüksek düzeyde sahip olanlar) genişletilmiş-alternatif veya klasik Likert tipi ölçeklere göre değişip değişmediği incelenmesi amaçlanmıştır.

Alt Amaçlar

1. Farklı örneklemelere uygulanan genişletilmiş-alternatif ve klasik Likert tipi ölçeklerden elde edilen geçerlilik ve güvenirlik kanıtları nasıldır?

- a) Ölçeklerin madde faktör yükleri arasında nasıl bir farklılık vardır?
- b) Uyum indeksleri arasında nasıl bir farklılık vardır?
- c) Güvenirlik katsayıları arasında nasıl bir farklılık vardır?

2. a) Sadece klasik Likert tipi ölçek formunu alan öğrencilerin puanlarına ait elde edilen geçerlilik ve güvenirlik kanıtları, çocuk ve ergen grubuna göre farklılaşmakta mıdır?

b) Sadece genişletilmiş-alternatif Likert tipi ölçek formunu alan öğrencilerin puanlarına ait elde edilen geçerlilik ve güvenirlik kanıtları, çocuk ve ergen grubuna göre farklılaşmakta mıdır?

3. Aynı grupta yer alan öğrencilerin, genişletilmiş-alternatif Likert tipi ölçeklerden elde edilen toplam puan ortalamaları ile klasik Likert tipi ölçeklerden elde edilen toplam puan ortalamaları farklılaşmakta mıdır?

4. Öğrenciler klasik Likert formda ve genişletilmiş-alternatif formda yer alan ifadeleri nasıl yorumlamaktadır?

Yöntem

Araştırma Modeli

Orijinalinde klasik Likert tipi ölçek olarak derecelendirilen çocuklar için sınav kaygısı ölçeğinin genişletilmiş-alternatif Likert tipi ölçeklere çevrilerek psikometrik özellikler bakımından karşılaştırılmasının amaçlandığı çalışma, karma modellerden açıklayıcı desende tasarlanmıştır. Bu desen iki aşamalıdır. Öncelikle nicel veriler toplanır ve analiz edilir. Daha sonra nicel verileri açıklamak ve ayrıntılı olarak incelemek için nitel veriler toplanıp analiz edilir (Creswell & PlanoClark, 2007). Bu bağlamda, öncelikle sınav kaygısı ölçeği öğrencilere uygulanıp analiz edilmiş ardından 20 öğrenci ile farklı yanıt seçeneklerini nasıl yorumladıklarını incelemek için sesli düşünme yaklaşımı ile bireysel görüşmeler yapılmıştır.

Çalışma Grubu

Çocuklar için sınav kaygısı ölçeği, 8-12 yaş aralığındaki öğrencilerin sınav kaygı düzeylerine ilişkin bilgi veren ölçme aracı olup hem yurt dışında hem de ülkemizde kullanılmaktadır. Bu çalışmada, evrene genelleme amacı taşımadığından öğrenciler seçkisiz olmayan örnekleme yöntemlerinden; derinlemesine araştırma yapabilmek ve çalışmanın amacına uygun durumların seçilmesi amacıyla amaçlı örnekleme yöntemi (Yıldırım & Şimşek, 2011) ve çalışmada zaman, para, emek, uygulanabilirlik vb. sınırlılıklardan dolayı kolay erişilebilir örnekleme yöntemleri kullanılarak Gaziantep ve Kilis illerinden Millî Eğitim Bakanlığına bağlı okullardan seçilmiştir. Çalışmanın nicel aşamasına çocuklar için sınav kaygısı ölçeği, 2023-2024 eğitim yılının birinci döneminde Gaziantep ve Kilis illerinden Milli Eğitim Bakanlığına bağlı okullardan 3. 4. 5. ve 6. sınıf düzeyinde öğrenim gören çalışmaya katılmaya gönüllü 892 öğrenciye uygulanmıştır. Ardından çalışmanın nitel aşamasında, 3. 4. 5. ve 6. sınıf düzeylerinde öğrenim gören katılmaya gönüllü, demografik özellikler açısından dengeli dağılım gösteren ve okuma-anlama becerisi yüksek 20 öğrenci ile bireysel görüşmeler yapılmıştır. Çalışmaya 445 kız 467 erkek olmak üzere toplam 912 öğrenci katılmış olup katılım sağlayan öğrencilerin 598'i çocuk grubunu 313'ü ergen grubunu oluşturmaktadır.

Veri Toplama Araçları

Bu çalışmada, veri toplama aracı olarak Çocuklar için Sınav Kaygısı Ölçeği kullanılmıştır. Bu çalışmada kullanılmak üzere ölçeği geliştiren yazarlardan izin alınmıştır. Ölçek hakkında bilgiler aşağıda sunulmuştur.

Çocuklar için Sınav Kaygısı Ölçeği (ÇKSÖ)

Wren ve Benson (2004) tarafından geliştirilen “Çocuklar İçin Sınav Kaygısı Ölçeği” ilkökul ve ortaokul çağındaki 8-12 yaş aralığındaki çocukların sınav kaygı düzeylerini ölçmeye yönelik olarak geliştirilen 30 maddeden oluşan, cevaplama süresi 5-12 dakika süren dörtlü Likert tipi (1 = Hemen Hemen Hiçbir Zaman, 2 = Bazen, 3= Çoğu Zaman, 4 = Hemen Hemen Her Zaman) bir derecelendirmeye sahip bir ölçme aracıdır. Öğrencilerin cevapları doğrultusunda ölçekten alınabilecek puan minimum 30 ve maksimum 120 puandır. 30–60 puan düşük, 60–90 puan orta ve 90–120 puan yüksek düzeyde sınav kaygısına işaret etmektedir (Wren & Benson, 2004). ÇSKÖ için güvenirlik kestirimi 0.92 olup alt boyutlarında ise Düşünceler, Görev Dışı Davranışlar ve Otonom Tepkiler için Cronbach Alfa değerleri sırasıyla .89, .78, ve .85 olarak hesaplanmıştır. Analizler, tüm madde-faktör yüklerinin istatistiksel olarak anlamlı olduğunu ve 0.37 ile 0.76 arasında değiştiğini (yarısından fazlası /0.60) göstermiştir (Wren & Benson, 2004). Türkiye’de Bulgan ve diğerleri (2017) tarafından yapılan ölçek uyarlama çalışmasında tutarlılık anlamında güvenirlik için iç tutarlılığı Cronbach Alfa katsayısı ile hesaplanmıştır. Ölçeğin bütününe ilişkin Cronbach Alfa değeri ise .88 olup alt boyutlarında ise Düşünceler, Görev Dışı Davranışlar ve Otonom Tepkiler için Cronbach Alfa değerleri sırasıyla .82, .70, ve .75 olarak bulunmuştur. Ölçeğe ait geçerlilik kanıtları ise yapılan analizler sonucunda 30 maddeden ve üç faktörden oluşan ölçeğin model uyumunu ki kare değeri ve $\chi^2/sd = 3.97$ dikkate alındığında χ^2/sd değerinin 5’in altında olduğu, anlamlı olduğu görülmüştür. Model uyum indeksleri ise RMSEA = .05, SRMR = .05, CFI = .97, GFI = .92 ve AGFI = .90 olarak hesaplanmış ölçeğin model uyum indekslerinin kabul edilebilir ve iyi düzeyde olduğu ve ölçeğin yapı geçerliğine sahip olduğu söylenebilir (Aydın ve Bulgan, 2017). Aşağıda klasik formatta uygulanan örnek bir madde (Aydın & Bulgan, 2017) ardından genişletilmiş hali örnek olarak sunulmuştur.

Klasik form:

	Neredeyse Hiçbir Zaman	Bazen	Çoğu Zaman	Neredeyse Her Zaman
Sınav olurken...				
1. Geçip geçmeyeceğimi merak ederim.	1	2	3	4

Genişletilmiş form:

Soru 1:

- A. Sınav olurken geçip geçmeyeceğimi neredeyse hiçbir zaman merak etmem.
- B. Sınav olurken geçip geçmeyeceğimi bazen merak ederim.
- C. Sınav olurken geçip geçmeyeceğimi çoğu zaman merak ederim.
- D. Sınav olurken geçip geçmeyeceğimi neredeyse her zaman merak ederim.

Nitel veriler için 20 katılımcı ile 40 dakikalık bireysel görüşmeler yapılmış olup sesli düşünme yaklaşımı kullanılmıştır. Klasik Likert tipi ölçekte yer alan ifadeleri ve genişletilmiş-alternatif Likert tipi ölçekte yer alan farklı yanıt seçenekleri; öğrencilere tek tek sorulup yanıtları kayıt altına alınarak veriler toplanmıştır. Nitel veriler için farklı bir veri toplama aracı kullanılmamıştır.

Etik Kurul Kararı

Ankara Üniversitesi Rektörlüğü Etik Kurul Başkanlığı 11.07.2023 tarihli ve 8543427050.04.04/990660 sayılı yazısı.

Milli Eğitim Bakanlığı Strateji Geliştirme Başkanlığı 23.10.2023 tarihli ve E-49614598-605.01-87840915 sayılı yazısı.

Araştırmanın Bilimsel Niteliği

Araştırmanın nicel boyutunda ölçme araçlarının güvenilirliğini değerlendirmek amacıyla Cronbach alfa katsayıları hesaplanmış, yapı geçerliğine ilişkin kanıt elde etmek amacıyla DFA gerçekleştirilmiştir. Araştırmanın nitel boyutunda ise inandırıcılığı artırmak amacıyla veri analiz sürecinde ilk yazar tarafından gerçekleştirilen kodlamalar ikinci yazar tarafından gözden geçirilmiş ve kodlamalar arasındaki tutarlılık değerlendirilmiştir. Ayrıca, bulguların katılımcı görüşlerini doğru biçimde yansıtmalarını sağlamak amacıyla doğrudan katılımcı ifadelerinden alıntılara yer verilmiştir.

Verilerin Analizi

Verilerin analizinde ölçekte yer alan her maddenin olumlu ya da olumsuz olmasına bakılmaksızın 1'den 4'e doğru cevaplayıcıların yanıtları puanlanmıştır. Elde edilen veriler SPSS ve Mplus programları kullanılarak araştırma kapsamında yer alan araştırma sorularına uygun yöntemler kullanılarak çözümlenmiştir. Analizler öncesi testlerin varsayımları test edilmiştir. Analizlerde manidarlık düzeyi .05 olarak alınmıştır. Analizlere başlamadan önce klasik Likert tipi formun, genişletilmiş-alternatif formun ve hem klasik hem genişletilmiş-alternatif formun uygulandığı gruplardan elde edilen verilere ilişkin betimsel istatistikler, kayıp veriler, uç veriler incelenmiş, normal dağılım varsayımı kontrol edilmiştir. Birinci ve ikinci araştırma sorusu kapsamında, genişletilmiş-alternatif ve klasik Likert tipi ölçeklerden elde edilen her bir formun güvenilirliğini Cronbach Alfa iç tutarlılık katsayısı hesaplanarak belirlenmiştir. Formlardan elde edilen güvenilirlik katsayılarının birbirinden anlamlı olarak farklılaşp farklılaşmadıkları Fisher'in Z Testi ile incelenmiştir (Akhun, 1984). Bu çalışmada farklı grupların karşılaştırılması, hataların rastgele dağıldığı varsayımına dayanmaktadır; bu durum araştırmanın bir sınırlılığı olarak kabul edilebilir. Geçerlik kanıtları için DFA yapılmış, genişletilmiş-alternatif ve klasik Likert tipi ölçeklerden elde edilen model veri uyum değerleri ve faktör yükleri karşılaştırılmıştır. Madde faktör yükleri arasındaki farklılık bağımsız örneklem t testi ile karşılaştırılmıştır. Üçüncü alt amaç kapsamında, hem klasik hem genişletilmiş-alternatif formun uygulandığı aynı grupta yer alan cevaplayıcıların

toplam puanlarının tüm ölçek için ve ölçeğin alt boyutlarından elde edilen toplam puanlar arasındaki genişletilmiş-alternatif Likert tipi ölçekleri alma durumlarına göre farklılaşma olup olmadığının belirlenmesi amacıyla bağımsız örneklem t testi kullanılmıştır. Aynı gruptan genişletilmiş-alternatif Likert tipi ölçekte aynı toplam puanı almış farklı kişilere ait profiller (maddelere ilişkin verilen tepkilerin dağılımı) ile klasik Likert tipi ölçekte aynı toplam puanı almış farklı kişilere ait profiller madde ve alt ölçek bazında betimsel olarak incelenerek farklılıklar sunulmuştur. Dördüncü alt amaç kapsamında, öğrencilerin klasik formda yer alan ifadeleri ve genişletilmiş-alternatif formda yer alan farklı yanıt seçeneklerini nasıl yorumladıklarını tespit etmek için öğrenciler ile yapılan görüşme kayıtları deşifre edilmiş ardından içerik analizi yapılarak sonuçlar tema, kategori ve kodlar ile sunulmuş, öğrencilerin görüşlerine ilişkin alıntılar verilmiştir.

Bulgular

Çalışmanın ilk amacı farklı gruplara uygulanan genişletilmiş-alternatif Likert tipi ölçekler ve klasik Likert tipi ölçeklerden elde edilen geçerlilik, güvenilirlik kanıtlarının incelenmesi ve güvenilirlik katsayıları, uyum indeksleri ve madde faktör yükleri arasında fark olup olmadığının tespit edilmesidir. Birinci araştırma sorusu kapsamında farklı gruplara uygulanan genişletilmiş-alternatif ve klasik Likert tipi ölçeklerden elde edilen her bir formun güvenilirliği Cronbach Alfa iç tutarlılık katsayısı hesaplanarak belirlenmiş olup güvenilirlik katsayılarının birbirinden manidar bir biçimde farklılaşıp farklılaşmadıkları ise Fisher'in Z Testi ile test edilmiş olup genişletilmiş-alternatif ve klasik Likert tipi ölçek formlarına ait Cronbach Alfa güvenilirlik katsayıları Ek 1'de verilmiştir. Elde edilen güvenilirlik katsayılarının birbirinden manidar bir biçimde farklılaşıp farklılaşmadıklarına ilişkin Fisher'in Z Testi ile sonucu Tablo 1'de verilmiştir.

Tablo 1

Formlara Ait Güvenirlik Sonuçlarının Karşılaştırılması

	Klasik Form	Genişletilmiş-Alternatif Form	Zrklasik	Zrgenişletilmiş	Z
Tüm ölçek	.88	.87	1.37	1.35	.30
Düşünceler	.83	.81	1.18	1.14	.64
Görev D.D.	.64	.61	.77	.74	.42
Otonom T.	.76	.78	1.00	1.05	-.70

Tablo 1'de görüldüğü üzere tüm z değerleri, 1.96'dan (Akhun, 1984) küçük olduğundan klasik ve genişletilmiş-alternatif Likert tipi ölçeklerin güvenilirlik katsayıları arasında manidar bir fark olmadığı görülmektedir. Diğer bir ifadeyle, klasik Likert tipi ölçeği genişletilmiş-alternatif forma dönüştürmek hem tüm ölçekten elde edilen güvenilirlik katsayısı üzerinde hem de ölçeğin alt boyutlarından elde edilen güvenilirlik katsayıları üzerinde bir fark yaratmamıştır. Geçerlik kanıtları için

doğrulamalı faktör analizi yapılmış, genişletilmiş-alternatif ve klasik Likert tipi ölçeklerden elde edilen model veri uyum değerleri ve faktör yükleri karşılaştırılmış, klasik Likert tipi ölçek ve genişletilmiş-alternatif Likert tipi ölçek formundan elde edilen verilerinin model uyum iyiliği doğrulamalı faktör analizi ile incelenmiş olup Tablo 2’de gösterilmiştir.

Tablo 2

Klasik Likert Tipi ve Genişletilmiş-Alternatif Likert Tipi Ölçek Formuna Ait Model Uyum İyiliği

Test Formu	χ^2	χ^2/sd	RMSEA	SRMR	CFI
Klasik Form	5507.05	12.65	.05	.06	.89
Alternatif Form	5627.08	12.9	.04	.05	.93

Uyum indekslerinden, CFI için ideal değer .90’dan büyük, RMSEA değerinin ise .08’den küçük olması gerekmektedir (Uyumaz & Çokluk, 2016). Ölçeklerden elde edilen uyum iyiliği değerleri incelendiğinde, χ^2/sd değerinin 5’in altında olmaması model uyumunun iyi olmadığını göstermektedir. Klasik forma ait RMSEA değeri .05 olup iyi uyum; genişletilmiş-alternatif forma ait RMSEA değeri .04 ise mükemmel uyum göstermektedir. SRMR değerleri bakımından klasik formun .06 ve genişletilmiş-alternatif forma ait .05 olduğu ve her ikisinin de iyi uyum gösterdiği görülmektedir. Klasik formdan elde edilen CFI değeri .89 olması nedeniyle iyi uyuma oldukça yakın olduğu görülmektedir. Uyum indekslerinden CFI, SRMR ve RMSEA genel olarak değerlendirildiğinde, doğrulamalı faktör analizi sonucunda üç faktörlü ölçeğin yapısının kabul edilebilir ve geçerli sonuçlar verdiği görülmektedir. Zhang ve Savelei (2016), çalışmada klasik ve genişletilmiş-alternatif formların model uyumlarının karşılaştırılmasında Ki-kare, CFI ve RMSEA değerlerini dikkate almış, genişletilmiş-alternatif formun uyum değerlerinin daha yüksek olduğu tespit edilmiştir. Bu çalışmada da ilgili bulguya paralel olarak formlar arasında küçük farkların olduğu ve genişletilmiş-alternatif formun daha iyi bir uyuma sahip olduğu tespit edilmiştir. Yani Zhang ve Savalei’nin bulgularıyla tutarlı olduğu görülmüştür. Orijinal formun uygulandığı gruptan elde edilen doğrulamalı faktör analizine ilişkin yol (path) diyagramları incelendiğinde genişletilmiş-alternatif forma ait tüm maddelerin, faktör yükleri .30’dan büyük olduğu ve üç faktörlü yapıyla iyi bir uyum gösterdiği; klasik formda ise madde 18 hariç tüm maddelerin, faktör yüklerinin .30’dan büyük olduğu ve üç faktörlü yapıyla iyi bir uyum gösterdiği görülmektedir. Kam ve diğerleri (2024), çalışmada genişletilmiş form için ölçekte yer alan maddelerin faktör yüklerini DFA ile incelemiş olup bu çalışmada da maddelerin faktör yükleri DFA ile incelenmiştir. Tüm maddelerin faktör yüklerinin oldukça yüksek olduğu ve orijinal yapı ile aynı faktöre yüklendiği tespit edilmiş olup iki farklı ölçekte yer alan maddelerin faktör yükleri bağımlı örneklem t testi ile karşılaştırılmış, sonuçlar Tablo 3’te sunulmuştur.

Tablo 3

Ölçekte Yer Alan Maddelerin Formlardaki Faktör Yüklerinin Bağımlı Örneklem t testi ile Karşılaştırılması (N: 30)

Formlar	M	ss	T	p
Klasik form ile Genişletilmiş-Alternatif formun faktör yükleri	.00	.71	.43	.66

Tablo 3'te görüldüğü üzere klasik Likert tipi ve genişletilmiş-alternatif Likert tipi ölçek formlarına ait faktör yükleri karşılaştırıldığında; t testi sonucu manidarlık düzeyi ($p=.66$) .05'ten büyük olduğundan ölçekte yer alan maddelerin klasik form ile genişletilmiş-alternatif formdaki faktör yükleri arasında anlamlı bir farklılık olmadığı ifade edilebilir ($t(29)=0.43, p=.66$). Analiz sonucunda genişletilmiş-alternatif formlar için model uyumlarının, klasik formlar için elde edilen model uyumlarından daha iyi uyum gösterdiği sonucuna ulaşılmıştır. Aynı zamanda faktörler altında toplanan madde faktör yükleri karşılaştırıldığında ise genişletilmiş-alternatif formdaki madde yüklerinin, klasik formdaki madde yüklerinden daha yüksek olduğu ve yapı geçerliliğinin daha iyi olduğu sonucuna ulaşılmıştır.

Araştırmanın ikinci alt amacı kapsamında farklı gruplardan genişletilmiş-alternatif Likert ve klasik Likert tipi ölçeklere ait elde edilen geçerlilik ve güvenilirlik kanıtlarının; çocuk ve ergen grubuna göre farklılaşıp farklılaşmadığı ayrı ayrı incelenmiştir. Elde edilen puanların güvenilirliği için genişletilmiş-alternatif Likert tipi forma ait Cronbach Alfa iç tutarlılık katsayısı incelendiğinde ergen grubundan elde edilen Cronbach Alfa değerinin ($\alpha=.91$) çocuk grubundan elde edilen Cronbach Alfa değerinden ($\alpha=.86$) daha yüksek olduğu; klasik forma ait Cronbach Alfa değerinin ergen grubunda elde edilen ($\alpha=.84$) ve çocuk grubundan elde edilen Cronbach Alfa değeri ($\alpha=.89$) olduğu bu durumda ergen grubunun Cronbach Alfa değerinin, çocuk grubundan daha düşük olduğu görülmekte olup sonuçlar Ek 2'de gösterilmiştir. Çocuk ve ergen grubundan elde edilen güvenilirlik katsayılarının birbirinden manidar bir biçimde farklılaşıp farklılaşmadıklarını tespit etmek için Fisher'in Z Testi kullanılmış olup formlara ait ergen ve çocuk grubu için Fisher'in Z Testi sonucu Tablo 4'te verilmiştir.

Tablo 4

Farklı Gruplardaki Çocuk ve Ergen Cevaplayıcılara Ait Güvenirlik Sonuçlarının Karşılaştırılması

	Klasik Form					Genişletilmiş-Alternatif Form				
	Ç	E	Zrç	Zre	Z	Ç	E	Zrç	Zre	Z
Tüm ölçek	89	84	1.42	1.25	1.63	.86	.91	1.29	1.52	-2.26*
Düşünce	82	83	1.17	1.19	-.15	.78	.87	1.05	1.33	-2.64*
Görev D.D.	68	55	.83	.62	2.01*	.61	.70	.71	.87	-1.54
Otonom	76	75	1.01	.98	.33	.75	.84	.98	.88	.95

* p<.05, Not: Ç=Çocuk Grubu, E=Ergen Grubu.

Tablo 4'te görüldüğü gibi sadece genişletilmiş-alternatif Likert tipi ölçek formunu yanıtlayan cevaplayıcıların görev dışı ve otonom tepkiler alt boyutlarında z değerleri, tablo değeri olan 1.96'dan küçük olduğundan çocuk ve ergen gruplarında ölçeklerin güvenilirlik katsayıları arasında manidar bir fark olmadığı görülmektedir. Diğer taraftan tüm ölçek ve düşünceler boyutuna ait güvenilirlik değerleri karşılaştırıldığında, çocuk ve ergen gruplarında ölçeklerin güvenilirlik katsayıları arasında manidar bir fark olduğu ifade edilebilir. Sadece klasik Likert tipi ölçek formunu yanıtlayan cevaplayıcıların Fisher z değerleri incelendiğinde ise düşünceler ve otonom tepkiler alt boyutlarında çocuk ve ergen gruplarında ölçeklerin güvenilirlik katsayıları arasında manidar bir fark olmadığı ancak tüm ölçek ve görev dışı tepkiler boyutuna ait güvenilirlik değerleri karşılaştırıldığında, çocuk ve ergen gruplarında ölçeklerin güvenilirlik katsayıları arasında manidar bir fark olduğu ifade edilebilir. Geçerlik kanıtları için çocuk ve ergen grubuna ait genişletilmiş-alternatif formdan elde edilen veriler için yapılan doğrulayıcı faktör analizi sonucunda model uyum iyiliği değerleri Tablo 5'te gösterilmiştir.

Tablo 5

Farklı Gruplardaki Çocuk ve Ergen Cevaplayıcılara Ait Genişletilmiş-Alternatif Form Uyum İndeksleri

Uyum İndeksleri		χ^2	χ^2/sd	RMSEA	SRMR	CFI
Genişletilmiş-Alternatif Form	Çocuk Grubu	574.08	1.42	.04	.06	.93
	Ergen Grubu	549.04	1.36	.04	.07	.93
Klasik Form	Çocuk Grubu	758,48	1.88	.04	.07	.90
	Ergen Grubu	531,92	1.32	.03	.09	.87

Genişletilmiş-alternatif Likert tipi forma ait uyum iyiliği değerleri incelendiğinde, RMSEA değeri hem çocuk grubu için (RMSEA=.04) hem de ergen grubu (RMSEA=.04) içinde mükemmel uyum gösterdiği görülmektedir. Ölçek formuna ait SRMR değerleri incelendiğinde hem çocuk grubuna hem ergen grubuna ait SRMR değerinin iyi uyum gösterdiği görülmektedir. Genişletilmiş- alternatif ölçek formuna ait CFI değerleri incelendiğinde ise çocuk grubunda ve ergen grubunda (CFI=.93) iyi uyuma karşılık geldiği görülmektedir. Uyum indeksleri genel olarak değerlendirildiğinde, çocuk ve ergen gruplarından elde edilen verilerin doğrulayıcı faktör analizi ile analizi sonucunda üç faktörlü ölçeğin yapısının kabul edilebilir ve geçerli sonuçlar verdiği görülmektedir. Klasik Likert tipi ölçek formunun çocuk ve ergen gruplarından elde edilen uyum iyiliği değerleri incelendiğinde ise, RMSEA değerinin çocuk grubunda (RMSEA=.04) ve ergen grubunda (RMSEA=.03) mükemmel uyum gösterdiği görülmektedir. Klasik Likert tipi ölçek formuna ait SRMR değerleri çocuk grubunda (SRMR=.07) ve ergen grubuna (SRMR=.09) iyi uyum gösterdiği görülmektedir. CFI değeri ise çocuk grubunda (CFI=.90) iyi uyuma karşılık geldiği; ergen grubunda (CFI=.87) iyi uyuma karşılık geldiği görülmektedir. Uyum indeksleri genel olarak değerlendirildiğinde, çocuk ve ergen gruplarından elde edilen verilerin doğrulayıcı faktör analizi ile analizi sonucunda üç faktörlü ölçeğin yapısının kabul edilebilir ve geçerli sonuçlar verdiği görülmektedir. Klasik form uyum değerleri bakımından incelendiğinde, çocuk grubundaki uyum değerlerinin ergen grubuna göre daha iyi sonuçlar verdiği görülmektedir. Bazı uyum değerleri çocuk grubunda bazıları ise ergen grubunda daha iyi uyuma işaret ettiğinden net karşılaştırma yapılamamıştır. Ölçeklerde yer alan maddelerin faktör yükleri bağımsız örneklem t testi ile karşılaştırılmış, sonuçlar Tablo 6'da verilmiştir.

Tablo 6

Farklı Gruplardaki Çocuk ve Ergen Cevaplayıcılarda Yer Alan Maddelerin Faktör Yüklerinin Karşılaştırılması (N: 30)

	M	ss	T	p
Genişletilmiş Form Çocuk-Ergen Faktör Yükleri	-.06	.08	-3.94	.001
Klasik Form Çocuk-Ergen Faktör Yükleri	-.03	.11	-1.48	.14

Tablo 6'da görüldüğü üzere, genişletilmiş-alternatif Likert tipi ölçek formunun çocuk ve ergen grubuna ait faktör yükleri karşılaştırıldığında; ölçekte yer alan maddelerin çocuk grubu ile ergen grubunun faktör yükleri arasında anlamlı bir farklılık olduğu görülmektedir ($t(29)=-3.94$, $p<.05$). Elde edilen etki büyüklüğü (Cohen $d=0.72$) bu farkın orta-yüksek düzeyde olduğunu ve yaş grubunun faktör yükleri üzerinde belirgin bir etkisi bulunduğunu göstermektedir. Klasik Likert tipi ölçek formunun çocuk ve ergen grubuna ait faktör yükleri karşılaştırıldığında; t değerinin manidarlık düzeyinin ($p=.14$) .05'ten büyük olduğu, yani ölçekte yer alan

maddelerin çocuk grubu ile ergen grubunun faktör yükleri arasında anlamlı bir farklılık olmadığı görülmektedir ($t(29)=-1.48, p>.05$). Çocuk ve ergen grubundan elde edilen maddelerin faktör yüklerine bakıldığında düşünceler alt boyutunda çoğu maddenin yaklaşık yarısının çocuk grubunda, ergen grubundan daha yüksek olduğu görülmektedir. Çocuk ve ergen gruplarında formlara ilişkin faktör yük ortalamaları ölçek ve alt boyutlar kapsamında Tablo 7’de gösterilmiştir.

Tablo 7

Farklı Gruplardaki Çocuk ve Ergen Cevaplayıcılardan Formlara Ait Faktör Yüklerinin Ortalaması

	Çocuk Grubu		Ergen Grubu	
	Klasik Form	Genişletilmiş Alternatif Form	Klasik Form	Genişletilmiş Alternatif Form
Tüm ölçek	.56	.52	.53	.58
Düşünceler	.57	.54	.57	.61
Görev D.D.	.53	.47	.44	.49
Otonom T.	.58	.55	.55	.64

Tablo 7’de görüldüğü üzere klasik formun ölçek ve alt boyutlar bazında çocuk grubunda faktör yükleri daha yüksek iken genişletilmiş-alternatif formun ölçek ve alt boyutlarda faktör yük ortalamalarının ergen grubunda daha yüksektir. Bu durumda ergen grubu için ölçeği genişletmenin işe yaradığı yorumu yapılabilir. Faktör yükleri madde bazında iki formda da incelendiğinde, düşünceler alt boyutunda yer alan madde 1 ve madde 29’un görev dışı davranışlar alt boyutunda ise madde 7 ve madde 18’in iki formda da çocuk grubunda düşük olduğu görülmüştür.

Araştırmanın üçüncü alt amacı kapsamında aynı grupta yer alan cevaplayıcıların toplam puanlarının genişletilmiş-alternatif Likert tipi ölçekleri alma durumlarına göre farklılaşıp farklılaşmadığı bağımsız örneklem t testi ile analiz edilmiş olup analiz sonucu Tablo 8’de verilmiştir.

Tablo 8

Ölçek Toplam Puanlarının Farklı Ölçek Alma Durumuna Göre Karşılaştırılması

Ölçek Formu	N	M	ss	t	p
Alternatif Form	97	68.38	16.14	-2.86	.00
Klasik Form	97	70.55	16.84		

Tablo 8’de görüldüğü gibi hem klasik hem de genişletilmiş-alternatif Likert tipi ölçek formunun uygulandığı gruptaki bireylerin toplam puanları, ölçek alma durumlarına göre ilişkili örneklem t testi ile incelenmiştir. Analiz sonucunda, t değerinin manidar olması ($t(96) = -2.87, p < .05$) nedeniyle öğrencilerin ölçek toplam

puanlarının alınan ölçek türüne göre genişletilmiş-alternatif formun lehine farklılaştığı yorumu yapılabilir. Elde edilen etki büyüklüğü ($d= 0.29$), farkın küçük düzeyde olduğunu ve ölçek formunun toplam puanlar üzerindeki etkisinin sınırlı olduğunu göstermektedir.

Araştırmanın dördüncü alt amacı kapsamında cevaplayıcıların maddelere verdiği yanıtlar testin amacına uygun olan ve testin amacına uygun olmayan biçimde tema, kategori ve kodlar ile sunulurken farklı yanıt seçeneklerini nasıl yorumladıkları incelenmiştir. Klasik formda yer alan maddelere ilişkin cevaplayıcıların verdiği yanıtlar testin amacına uygun olan ve testin amacına uygun olmayan yanıtlar olarak incelenmiş ve maddelere ilişkin yorumları üç tema altında beş adet kategori ve 13 adet Kod oluşturulmuş olup cevaplayıcıdan alınan alıntılar Ek 3'te sunulmuştur. Ek 3'te kişisel özellikler kategorisi incelendiğinde, kopya çekmek kodu altında öğrencilerin ölçekte yer alan 'Sınav olurken arkadaşlarıma bakarım.' ve 'Sınav olurken etrafımda olup bitenlere bakarım.' ifadelerine verdikleri cevaplar incelendiğinde öğrencilerin amaca uygun olmayan yanıtlamalar yaptıkları da dikkat çekmiştir. Zaman yönetimi kapsamında yanıtlayan öğrencilerin testin amacına uygun yanıtlar verdikleri söylenebilir. Öğrencilerin ölçekte yer alan 'Sınav olurken yanlış bir şey yapmaktan korkarım', 'Sınav olurken başka şeyler hakkında düşünürüm.' ve 'Sınav olurken dalıp giderim.' ifadelerini dikkat koduyla açıkladıkları ve testin amacına uygun yanıtlama yapmadıkları görülmüştür.

Kişiyeye ait özellikler temasına ait duyuşsal özellikler kategorisinin ilk kodu "sınav korkusu/kaygısı" olup öğrencilerin verdikleri yanıtlara dikkat edildiğinde, testin amacına uygun yanıt verdikleri görülmüştür. Bu kategorinin ikinci kodu "heyecan" olup öğrencilerin verdikleri yanıtların testin amacına uygun olduğu görülmüştür. Bu kategorinin üçüncü kodu "sıkılganlık" olup öğrencilerin verdikleri yanıtlara bakıldığında, testin amacı dışında bir amaca hizmet ettiği söylenebilir. Sınava ait özellikler temasında öğrencilerin sınavın zorluğu kavramını farklı yorumladıkları, sınav zorluk düzeyini kendi başarıları doğrultusunda şekillendirdiklerini, bazı öğrencilerin sınav kavramını ders ile ilişkilendirdiği ve sadece bir dersi dikkate alarak yanıtladıkları görülmüş olup sonraki çalışmalar da maddelerin belirsizliklerinin giderilerek standartlaştırılması gerektiği düşünülmektedir. Aileye ait özellikler kapsamında öğrencilerin sınava yönelik kaygılarında ailenin tutumu ile ilgili olabileceği görülmektedir. Öğrencilerden alınan yanıtlar doğrultusunda bu tema da maddenin oldukça açık ve net olduğu, öğrencilerin kolaylıkla anladığı, farklı anlamlar yüklemelikleri görülmüştür. Genişletilmiş formda yer alan maddelere ilişkin öğrencilerin verdiği yanıtlar testin amacına uygun olan ve testin amacına uygun olmayan yanıtlar olarak incelenmiş ve maddelere ilişkin yorumları üç tema altında beş adet kategori ve 13 adet kod oluşturulmuş olup Ek 4'te sunulmuştur.

Kişiyeye ait özellikler teması kişisel özellikler; bilişsel özellikler ve duyuşsal özellikler olmak üzere üç kategoriden ve 10 koddan oluşmaktadır. İlk olarak kişisel özellikler kategorisi incelenmiş olup bu öğrencilerin amaca uygun olmayan

yanıtlamalar yaptıkları dikkat çekmiştir. Öğrencilerin sınav olurken cevapları hatırlamak için düşündükleri ölçekte yer alan sınav olurken kalemiyle oynadıkları, başının ağrıdığını, ayaklarını sallamaya başladıkları ifadeleri bilgiyi geri getirme kodu altında toplandığı görülmüştür. Kişiye ait özellikler temasına altında toplanan bulgular incelendiğinde, öğrencilerin maddelerin çoğunu amacına uygun yanıtladıkları, öğrenci yanıtlarının birbirine benzer olduğu, yanıt çeşitliliğinin klasik formdan alınan yanıtlarla karşılaştırıldığında daha az olduğu ve maddeler için alınan cevapların belli kategorilerde ağırlıkla toplandığı görülmüştür. Bu durumda genişletilmiş-alternatif formda yer alan maddelerin, klasik forma oranla daha anlaşılır ve net olduğu, öğrencilerin maddeleri farklı yorumlasalar dahi bu farklılığın az olduğu görülmüştür. Sınavın güçlüğüne ilişkin bir kategori, sınavın zorluğu ve sınavın kolaylığı olarak da iki kod oluşturulduğu görülmektedir. Bu kodlarda sınavın kolaylığı kapsamında iki numaralı öğrenci, sınavı hızlı bitirmeye çalışırım ifadesine öğrencinin verdiği cevap alıntı kısmında yer verilmiştir. Sınavın zorluğu kodu kapsamında; bir numaralı öğrencinin cevapları hatırlamakta zorlanırım ifadesine; 7 numaralı öğrencinin cevaplarının doğru olup olmadığını merak ederim ifadesine; 10 numaralı öğrencinin ise sınav olurken başarısız olmaktan korkarım ifadesine verdikleri yanıtlar alıntı kısmında yer verilmiştir.

Genişletilmiş formun uygulandığı öğrencilerin cevapları dikkate alındığında bazı ifadeler ölçüğün amacına uygun olmayan yanıtlar verdikleri, testin amacı dışında farklı bir amaca hizmet ettiği görülmüştür. Öğrencilerin ifadeleri dikkate alındığında genellikle her zaman ya da hiçbir zaman zaman olarak işaretledikleri “kalemimle oynarım, karnımda tuhaf ağrı olur, saati kontrol ederim, yerimde hareketsiz durmakta zorlanırım, daha fazla çalışmalıydım diye düşünürüm, tuvalete gitme ihtiyacı hissederim” gibi ifadelerle ait işaretlemelerinde zaman unsuruna bağlı yanıtladıkları; bazen ve çoğu zaman seçeneklerini işaretleme eğilimlerinin fazla olduğu etrafıma çevremde olup bitenlere bakarım, tedirgin hissederim, cevaplarımın doğru olup olmadığını merak ederim, sınavımı hızlı bitirmeye çalışırım ifadelerle ait işaretlemelerinde ise eylem unsuruna bağlı yanıtladıkları görülmüştür. Öğrencilerden ölçekte yer alan maddelerle ait alınan yanıtlarda ‘Sınav olurken tuvalete gitme ihtiyacı hissederim.’ maddesini yanıtlarken öğrencilerin tamamının zamana; ‘Sınav olurken dalıp giderim.’ maddesini yanıtlarken ise tamamının eyleme dönük yanıtlamada bulundukları dikkat çekmiştir.

Tartışma, Sonuç ve Öneriler

Bu çalışmada klasik Likert tipi ölçeklerin, genişletilmiş-alternatif Likert tipi ölçeklere çevrilerek psikometrik özellikler bakımından karşılaştırılması, genişletilmiş-alternatif Likert tipi ölçek formunun farklı yaş gruplarında (çocuk-ergen grup) olmanın geçerlik ile güvenilirliğe etkisi ve ölçek toplam puanı genişletilmiş alternatif veya klasik Likert tipi ölçeklere göre değişip değişmediği incelenmiştir. Bu doğrultuda orijinalinde klasik Likert tipi ölçek formunda olan çocuklar için sınav kaygısı ölçeği; uzman görüşleri ve çocuklardan sesli düşünme yöntemi ile alınan cevaplar doğrultusunda genişletilmiş formata çevrilmiş olup 3. 4. 5. ve 6. Sınıf

öğrencilerine uygulanmıştır. Aşağıda bu araştırmadan elde edilen sonuçlar raporlanmıştır.

Çalışmanın ilk alt amacı doğrultusunda farklı gruplardan genişletilmiş-alternatif Likert tipi ölçeklere ait elde edilen güvenilirlik kanıtları ve klasik Likert tipi ölçeklerden elde edilen güvenilirlik kanıtlarının iki formda da yüksek olduğu ancak iki form arasında istatistiksel açıdan manidar bir farklılaşma olmadığı görülmüştür. Zhang ve diğerleri (2023), çalışmasında Rosenberg Benlik Saygısı ölçeğine ait hem klasik hem de genişletilmiş-alternatif Likert tipi ölçek formu yetişkin grubunda uygulamış olup klasik forma ait Cronbach Alfa değeri ile genişletilmiş-alternatif forma ait Cronbach Alfa değeri birbirine yakın olarak hesaplanmıştır. Klasik Likert tipi ölçeğin, genişletilmiş-alternatif Likert tipi ölçek formuna dönüştürmek hem tüm ölçekten elde edilen güvenilirlik katsayısı üzerinde hem de ölçeğin alt boyutlarından elde edilen güvenilirlik katsayıları üzerinde bir fark yaratmadığı sonucuna ulaşılmıştır. İki formda dikkate alındığında görev dışı davranışlar alt boyutundan elde edilen değerlerin orijinal forma göre daha düşük olduğu bu durumun alt boyutta yer alan maddelerin açık ve net olmaması, çocukların okuduğunu anlama becerisinden, çalışma grubunda farklı kültürlerden katılan ve küçük yaş gruplarındaki öğrencilerin ifadelerde yer alan kavramları tam anlayamamasından kaynaklandığı düşünülmektedir. Örneğin; “Yüzümü sıcak basmaz.” ifadesinde “sıcak basmaz” deyimini yaş grubu küçük olan bireylerin ve farklı kültürlerden gelen öğrencilerin tam anlamadığı gözlemlenmiştir.

Genişletilmiş-alternatif Likert tipi ölçeklere ait elde edilen geçerlilik kanıtları ve klasik Likert tipi ölçeklerden elde edilen geçerlilik kanıtları incelendiğinde ise uyum indeksleri ve faktör yükleri karşılaştırılmış olup iki formda da tüm maddelerin manidar ve genişletilmiş-alternatif formun model uyum değerleri daha iyi olan bir model olabileceği sonucuna ulaşılmıştır. Bu çalışmada da ilgili bulguya paralel olarak formlar arasında küçük farkların olduğu ve genişletilmiş-alternatif formun daha iyi bir uyuma sahip olduğu tespit edilmiş olup Zhang ve Savalei'nin bulgularıyla tutarlı olduğu görülmüştür.

Genişletilmiş alternatif Likert tipi ölçeklere ait elde edilen geçerlilik kanıtları ve klasik Likert tipi ölçeklerden elde edilen geçerlilik kanıtları kapsamında DFA sonuçları incelendiğinde, ölçekte yer alan maddelere ait faktör yük değerlerinin ölçek ve alt boyutlarında yüksek olması nedeniyle iki ölçek formunun da yapı geçerliğine sahip olduğu ancak iki form arasında aralarında anlamlı fark olmadığı sonucuna ulaşılmıştır. Zhang ve Savalei (2016), çalışmasında Beck Depresyon ölçeği ve Rosenberg Benlik saygısı ölçeğini kullanmış olup DFA analizi ile genişletilmiş formattaki ölçeklerin, klasik Likert formdaki ölçeklere göre faktörleşme yapısının daha iyi olduğu görülmüştür. Bu çalışmada ise Çocuklar İçin Sınav Kaygısı ölçeği kullanılmış olup bulgulara bakıldığında iki ölçeğinde benzer faktörleşme yapısına sahip olduğu bu durumun kullanılan ölçme aracından ve ölçmenin uygulandığı kişilerden kaynaklanabileceği düşünülmüştür. Faktör yüklerinin ortalamalarına bakıldığında ve maddelerin klasik ve genişletilmiş- alternatif formdaki faktör yükleri

ayrı ayrı analiz edildiğinde, durumun alt boyutlara göre değişiklik gösterdiği görülmüştür.

Çalışmanın ikinci alt amacı doğrultusunda farklı gruplardan klasik Likert tipi ve genişletilmiş-alternatif Likert tipi ölçek formlarına ait elde edilen güvenilirlik kanıtları, çocuk ve ergen grubuna göre incelenmiş olup klasik formdan çocuk ve ergen grubuna ait elde edilen z değerleri dikkate alındığında görev dışı davranışlar alt boyutunda, genişletilmiş- alternatif formun ise tüm ölçek ve düşünceler alt boyutunda çocuk ve ergen grubundan elde edilen z değerlerine bakıldığında güvenilirlik katsayıları arasında manidar fark varken iki formda da otonom tepkiler alt boyutunda çocuk ve ergen grupları arasında manidar fark olmadığı görülmüştür. Görev dışı tepkiler alt boyutunda ise klasik formdan çocuk ve ergen grubuna ait güvenilirlik katsayıları arasında manidar fark varken genişletilmiş-alternatif formda manidar fark olmadığı görülmüş olup bu durumun maddelerin daha açık hale getirilmesi ve buna bağlı kafa karışıklığının azalmasından kaynaklanabileceği düşünülmüştür. Düşünceler alt boyutunda ise genişletilmiş-alternatif formda çocuk ve ergen grubu arasında manidar fark varken klasik formda manidar fark olmadığı görülmüştür. Düşünceler alt boyutu için genişletilmiş-alternatif formda ergen grubunun maddeleri ölçeğin anlamına daha uygun yanıtladıkları görülmüştür. Bu durum çocuk grubunun somut işlemler döneminde yer almasından dolayı soyut ifadeleri tam anlayamadıklarından kaynaklandığı düşünülmektedir. Son olarak otonom tepkiler alt boyutuna bakıldığında hem klasik hem de genişletilmiş-alternatif formdan çocuk ve ergen grubuna ait güvenilirlik katsayıları arasında manidar fark olmadığı görülmüştür. Bu çalışmada klasik form, çocuk grubunda nispeten daha güvenilirken ergen grubunda ise genişletilmiş-alternatif formun daha güvenilir olduğu görülmüştür. Bu durumun çalışma grubundan ve veri toplama aracından kaynaklanabileceği, genişletilmiş-alternatif formda yer alan maddelerin uzun olması ve okuma yükünü arttırması nedeniyle çocuk gruplarında ölçmeye karışan tesadüfi hatalardan daha fazla etkilendiği düşünülmektedir. Farklı gruplardan klasik Likert tipi ve genişletilmiş-alternatif Likert tipi ölçeklere ait elde edilen geçerlilik kanıtları, çocuk ve ergen grubuna göre uyum değerleri incelendiğinde, çocuk grubunda genişletilmiş- alternatif formun, ergen grubunda ise klasik formun daha iyi uyum değerine sahip olduğu görülmektedir. Gruplara uygulanan formlara ilişkin faktör yükleri incelendiğinde ise tüm maddelerin faktör yük değerlerinin istatistiksel olarak anlamlı olması, faktör yük değerlerinin yüksek olması ve model veri uyum indeksi değerleri birlikte değerlendirildiğinde, hem klasik formun hem de genişletilmiş alternatif formun çocuk ve ergen gruplarından elde edilen puanların yapı geçerliğine sahip olduğu söylenebilir. Klasik ve genişletilmiş alternatif Likert tipi ölçeklere ait faktör yükleri çocuk ve ergen gruplarında karşılaştırıldığında, genişletilmiş-alternatif formda ergen grubu lehine anlamlı farklılık varken klasik formda anlamlı bir fark olmadığı görülmüştür.

Aynı grupta genişletilmiş-alternatif veya klasik Likert tipi ölçeklerden aynı bireylerden elde edilen toplam puanlar arasında manidar fark olduğu, klasik formdan alınan toplam puanların, genişletilmiş-alternatif formdan alınan toplam puanlardan

daha düşük olduğu sonucuna ulaşılmıştır. Bu durumun ölçekte yer alan maddeleri genişletmenin fark yarattığı, maddelerinin sıklığının daha net belirtilmesinden kaynaklandığı düşünülmektedir. Ayrıca klasik form ile genişletilmiş-alternatif formların uygulanması arasındaki zaman farkının az olmasının ise iki uygulama arasındaki zamanın az olmasının, bireylerin maddelere verdikleri yanıtları hatırlamalarından da kaynaklandığı düşünülmektedir.

Öğrencilerin ölçek puanı aynı olsa da maddelere farklı yanıtlar verdikleri gözlemlendiğinden, bireylerin profillerini ayırtmada ve ölçek maddelerinin amaca hizmet etme durumunu tespit etmede ölçeğin farklı formları yerine bireysel görüşmelerin daha etkili olduğu tespit edilmiştir. Zhang ve diğerleri (2019) çalışmasında klasik formun maddelerinin belirsiz ve ifadelerin muğlak olmasının dezavantaj olduğunu vurgulamış olup bu çalışmada öğrencilerden alınan yanıtlar doğrultusunda oluşturulan tema incelendiğinde klasik formda yer alan bazı maddelerin yeterince açık ve net olmadığı bu nedenle öğrencilerin maddelere farklı yanıtlar verdikleri görülmüştür. Bu durum ölçekleme sürecinden ziyade maddelerin açıklığı ve anlaşılabilirliğiyle ilgili olup hepsinin bir arada değerlendirilmesi gerektiğini de göstermektedir. Bu çalışmanın bulguları birlikte değerlendirildiğinde, klasik Likert tipi ölçek ile genişletilmiş-alternatif Likert tipi ölçek formunun güvenilirlik ve genel yapı geçerliği açısından benzer psikometrik özellikler sergilediği görülmektedir. Bu durum, genel ölçme amacıyla klasik Likert tipi formun kullanılmaya devam edilebileceğini göstermektedir. Ancak, yaş gruplarına göre elde edilen uyum indeksleri dikkate alındığında, çocuk grubunda genişletilmiş-alternatif formun daha iyi uyum göstermesi, bu yaş grubunda maddelerin daha açık ve somut ifade edilmesinin ölçme kalitesini artırabileceğine işaret etmektedir. Buna karşılık ergen grubunda klasik formun daha iyi uyum göstermesi, bu grubun daha soyut ve genel ifadeleri anlamlandırmada yeterli olduğunu düşündürmektedir. Öte yandan, aynı bireylerden elde edilen toplam puanlar arasında anlamlı farklılık bulunması ve klasik formdan elde edilen puanların daha düşük olması, madde formatının bireylerin yanıt verme davranışını etkilediğini ortaya koymaktadır. Bu bulgu, farklı formattaki ölçeklerden elde edilen puanların doğrudan karşılaştırılmaması gerektiğini de göstermektedir. Bireylerin toplam puanları benzer olsa dahi maddelere verdikleri yanıtların farklılaşması, özellikle bireysel düzeyde ayrıntılı değerlendirme yapılmak istendiğinde ölçek formlarının tek başına yeterli olmayabileceğini ortaya koymaktadır. Bu nedenle, bireysel profilin daha derinlemesine anlaşılmasının amaçlandığı durumlarda ölçek uygulamalarının bireysel görüşmeler gibi nitel yöntemlerle desteklenmesi önerilmektedir.

Bununla birlikte, elde edilen bulgular ölçek geliştirme ve uyarlama çalışmalarında maddelerin hedef yaş grubu tarafından nasıl anlaşıldığının sistematik biçimde incelenmesinin önemini açıkça ortaya koymaktadır. Özellikle sesli düşünme (thinking-aloud) yöntemiyle gerçekleştirilen ön uygulamaların, ölçek maddelerinin uygulanacağı yaş grubuna geçilmeden önce küçük bir örneklem üzerinde yürütülmesi önerilmektedir. Nitekim bu çalışmada da görüldüğü üzere, öğrencilerin ilgili ifadeleri okuduklarında zihinlerinde canlandırdıkları anlamların birbirinden farklılaşabildiği

anlaşılmaktadır. Bu durum, yalnızca dilsel açıklığın yeterli olmadığını; aynı zamanda bilişsel süreçlerin de dikkate alınması gerektiğini göstermektedir. Dolayısıyla, ölçek maddelerinin hedef kitle tarafından nasıl yorumlandığını ortaya koyan nitel ön çalışmaların, ölçme aracının geçerlik kanıtlarını güçlendirecek kritik bir adım olduğu düşünülmektedir.

Çalışma bulgularına dayalı olarak, ölçek uygulamalarında açıklama metnlerinin daha kısa ve anlaşılır biçimde düzenlenmesi önerilmektedir. Farklı derecelendirme türlerine (3'lü, 5'li, 7'li vb.), madde sayılarına (kısa–uzun form) ve yanıt kategorilerine göre psikometrik karşılaştırmaların yapılması gelecekteki çalışmalar için önem taşımaktadır. Elde edilen bulgular doğrultusunda, genişletilmiş-alternatif formun özellikle soyut düşünme becerisi gelişmiş gruplarda (ergenlerde) tercih edilmesi uygun görülürken, klasik formda bazı maddelerin farklı yorumlanabilmesi nedeniyle bu formu kullanacak araştırmacıların dikkatli olması gerekmektedir. Ayrıca, bireysel görüşmeler ve sesli düşünme yöntemi gibi nitel yaklaşımların, maddelerin anlaşılma biçimini ve yapı geçerliğini değerlendirmede önemli katkı sağladığı görülmüştür. Son olarak, çocuk ve ergen grupları arasında faktör yüklerinin farklılaşması nedeniyle, ileriki çalışmalarda ölçme değişmezliğinin test edilmesi önerilmektedir.

References

- Akhun, İ. (1984). İki korelasyon katsayısı arasındaki manidarlığın test edilmesi. *Ankara Üniversitesi Eğitim Fakültesi Dergisi*, 17(1), 1-7. <https://dergipark.org.tr/tr/pub/auedfd/article/615250>
- Alan, Ü. (2019). *Likert tipi ölçeklerin çocuklarla kullanımında yanıt kategori sayısının psikometrik özelliklere etkisi [The effect of the number of response categories on the psychometric properties of Likert-type scales used with children.]* (Thesis No. 582737). [Master's thesis, Hacettepe University]. Council of Higher Education Thesis Center. <https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp>
- Aydın, U., & Bulgan, G. (2017). Turkish adaptation of the test anxiety scale for children: Validity and reliability study. *Elementary Education Online*, 16(2), 887-899. <https://doi.org/10.17051/ilkonline.2017.304742>
- Başaran, A. (1997). *Likert Tipi Ölçek Hazırlama Kılavuzu [Guide to preparing likert-type scales]*. Turkish Psychological Association Publications.
- Borgers, N., & Hox, J. J. (2001). Item nonresponse in questionnaire research with children. *Journal of Official Statistics*, 17(2), 321-335. <https://www.joophox.net/publist/jos01.pdf>
- Börkan, B., Öztömür, G., Yılmaz, O., Çetintaş, Ş., Gülcan, B., & Özcan, M. (2017). Rehberlik ve araştırma merkezlerinde psikolojik test uygulama süreçleri. *Turkish Psychological Counseling and Guidance Journal*, 7(48), 161-185. <https://izlik.org/JA99XF37CD>
- Cansız, R. O. (2007). *Sosyal Psikiyatride Kullanılan Psikolojik Testler [Psychological Tests Used in Social Psychiatry]*. [Unpublished course project]. Ankara University. <https://dspace.ankara.edu.tr/server/api/core/bitstreams/79d1ab9a-7c49-4119-b81a-13ca1b7a6ee9/content>
- Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests & measurement* (9th ed.). McGraw Hill.
- Gökdemir, F., & Yılmaz, T. (2023). Using, modifying, adapting, and developing likert-type scales. *Journal of Nursology*, 26(2), 148-160. <https://doi.org/10.5152/JANHS.2023.22260>
- Hills, P., & Argyle, M. (2002). The oxford happiness questionnaire: A compact scale for the measurement of psychological well-being. *Personality and Individual Differences*, 33(7), 1073-1082. [https://doi.org/10.1016/S0191-8869\(01\)00213-6](https://doi.org/10.1016/S0191-8869(01)00213-6)
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science and Technology*, 7(4), 396-403. <https://doi.org/10.9734/BJAST/2015/14975>

- Kam, C.C.S., Cheng, E. H., & Cui, T. (2024). Measuring self-esteem with expanded format in a fraction of time: ESE-S and ESE-US. *Journal of Personality Assessment*, 106(2), 196-207. <https://doi.org/10.1080/00223891.2023.2259990>
- Kutanis, R., & Elçi, M. (2006). The compatibility of psychotechnical tests with each other and individual characteristics: A validity analysis. *Journal of Academic Studies*, 1(2), 111-127. <https://dergipark.org.tr/en/pub/akademikincelemeler/article/19166>
- Royeen, C. B. (1985). Adaptation of likert scaling for use with children. *The Occupational Therapy Journal of Research*, 5(1), 59-69. <https://doi.org/10.1177/1539449285005001>
- Şahin, D. B., & Gülleroğlu, H. D. (2013). An examination of the psychometric properties of scales created using different item analysis techniques for item selection in likert-type scales. *Asian Journal of Education*, 1(2), 18-28. <https://doi.org/10.20897/apjes/16920>
- Turan, I., Şimşek, Ü., & Aslan, H. (2015). The use and analysis of the likert scale and likert-type questions in educational research. *Sakarya University Journal of Education*, 30(186), 186-203. <https://dergipark.org.tr/en/pub/sakaefd/article/134252>
- Tülübaş, G. (2009). *Psikolojik testlerde Angoff ve Sınır Grup Yöntemleri ile kesme puanlarının belirlenmesi [Determining Cut-off Scores in Psychological Tests Using the Angoff and Delineation Group Methods]. (Thesis No. 257642). [Master's thesis, Hacettepe University]. Council of Higher Education Thesis Center.* <https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp>
- Uyumaz, G., & Çokluk, Ö. (2016). Examining item ordering and rating differences in likert-type scales in terms of psychometric properties and respondent attitudes. *Journal of Theoretical Educational Science/Kuramsal Eğitim Bilim Dergisi*, 9(3), 400-425. <http://doi.org/10.5578/keg.10011>
- 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>
- Yıldırım, A. & Şimşek, H. (2011). *Sosyal bilimlerde nitel araştırma yöntemleri [Qualitative research methods in the social sciences]*. Seçkin Yayıncılık.
- Zhang, X., & Savalei, V. (2016). Improving the factor structure of psychological scales: The expanded format as an alternative to the likert scale format. *Educational and Psychological Measurement*, 76(3), 357-386. <https://doi.org/10.1177/0013164415596421>
- Zhang, X., Tse, W. W. Y., & Savalei, V. (2019). Improved properties of the big five inventory and the Rosenberg self-esteem scale in the expanded format relative

to the likert format. *Frontiers in Psychology*, 10(1286), 1-11.
<https://doi.org/10.3389/fpsyg.2019.01286>

Zhang, X., Zhou, L., & Savalei, V. (2023). Comparing the psychometric properties of a scale across three likert and three alternative formats: An application to the rosenberg self-esteem scale. *Educational and Psychological Measurement*, 83(4), 649-683. <https://doi.org/10.1177/00131644221111402>

Ethical and Author Declarations | Etik ve Yazar Beyanları**Graduate Thesis Information**

This article was derived from the Master's thesis entitled "Comparison of the Psychometric Properties of Extended Likert-Type Scales and Classical Likert-Type Scales in Psychological Tests", completed by Seda Gözlügöl under the supervision of Prof. Dr. Seher Yalçın at the Graduate School of Educational Sciences, Ankara University, in 2024 (Thesis No. 853769).

Authors' Contributions

This study was derived from the first author's Master's thesis conducted under the supervision of the second author. The first author played an active role in the planning of the study, data collection, data analysis, and preparation of the initial manuscript draft. The second author provided supervision and academic guidance in the design of the research, methodological decision-making, evaluation of the analyses, and revision of the manuscript. Both authors contributed to the writing of the article, reviewed the final version, and approved it for publication.

Conflict of Interest

No conflict of interest.

Ethical Approve

Letter dated July 11, 2023, No. 8543427050.04.04/990660, from the Chair of the Ethics Committee of the Rectorate of Ankara University.

Use of Artificial Intelligence

Declare whether any AI tools were used in the writing, analysis, or data processing stages, and specify the purpose if applicable. If no AI tools were used, this should also be clearly stated.

Lisansüstü Tez Bilgileri

Bu makale, Seda Gözlügöl' ün, Prof. Dr. Seher Yalçın Danışmanlığında Ankara Üniversitesi Eğitim Bilimleri Enstitüsünde 2024 yılında tamamlanan 853769 nolu "Psikolojik Testlerde Genişletilmiş Likert Tipi Ölçek İle Klasik Likert Tipi Ölçeklerin Psikometrik Özelliklerinin Karşılaştırılması" başlıklı Yüksek Lisans tezinden üretilmiştir.

Yazarların Katkı Düzeyleri

Bu çalışma, birinci yazarın ikinci yazar danışmanlığında yürüttüğü yüksek lisans tezinden üretilmiştir. Araştırmanın planlanması, veri toplama süreci, veri analizi ve ilk taslağın hazırlanması aşamalarında birinci yazar aktif rol almıştır. İkinci yazar araştırmanın tasarlanması, yöntemsel kararların verilmesi, analizlerin değerlendirilmesi ve çalışmanın gözden geçirilmesi süreçlerinde danışmanlık ve akademik rehberlik sağlamıştır. Her iki yazar da makalenin yazımına katkıda bulunmuş, son hâlini gözden geçirmiş ve onaylamıştır.

Çıkar Çatışması Beyanı

Herhangi bir çıkar çatışması bulunmamaktadır.

Etik Onay

Ankara Üniversitesi Rektörlüğü Etik Kurul Başkanlığı 11.07.2023 tarihli ve 8543427050.04.04/990660 sayılı yazısı.

Yapay Zeka Kullanımı

Araştırmanın yazım, analiz ve veri işleme aşamalarında herhangi bir yapay zekâ aracı kullanılmamıştır.

This study has been evaluated under double-blind peer review and verified to be free of plagiarism using iThenticate software.

Bu çalışma, çift taraflı kör hakemlik kapsamında değerlendirilmiş ve iThenticate yazılımı kullanılarak intihal içermediği teyit edilmiştir.

The articles published in our journal are made available as open access under the CC-BY-NCND license.

Dergimizde yayımlanan çalışmalar CC-BY-NCND lisansı ile açık erişim olarak yayımlanmaktadır.

Ethical disclosure | Etik bildirim: ebfd@ankara.edu.tr

Appendix

Appendix 1

Cronbach's Alpha Reliability Coefficients for the Extended-Alternative and Classic Likert Forms

	All Scale	Thoughts	Off-Duty Conduct	Autonomous Responses
Extended-Alternative Form	.87	.81	.61	.78
Classic Form	.88	.83	.64	.76

$**p < 0.01$

Appendix 2

Comparison of Cronbach's Alpha Internal Consistency Coefficients Obtained from Different Groups According to the Child and Adolescent Group

Form	Group	All Scale	Thoughts	Off-Duty Conduct	Autonomous Responses
Extended-Alternative Form	Child Group	.86	.78	.61	.75
	Adolescent	.91	.87	.70	.84
Classic Form	Child Group	.89	.82	.68	.76
	Adolescent	.84	.83	.55	.75

Appendix 3

Category and Code Table for Themes Related to Items in the Classic Form

Person-Related Characteristics			Features of the Exam	Family Characteristics
<i>Personal Characteristics</i>	<i>Cognitive Characteristics</i>	<i>Sensory Characteristics</i>	<i>The Difficulty of the Exam</i>	<i>The Family's Attitude</i>
Cheating: <i>Teacher, those who cheat here are constantly looking at the students behind them or in front of them, right</i> ? (S1*)	Note: <i>Yes, my attention wanders easily, so I get distracted thinking about other things</i> (S8).	Test Anxiety: If I fail, my mom will get really mad. I think they're asking if we're afraid of the test or not (S2).	Easy: It asks for our opinion on whether the exam will be difficult or easy. If the exam is easy, it doesn't worry me (S1).	Family Reaction: My mom can get angry sometimes if I get bad results, but she doesn't get angry often. I think she wants to see if our family will get angry or not (S5).
Time Management: I consider moving around too much during the exam to be a waste of time; I only focus on my paper (S10).	Forgetfulness: Sometimes I think this way because there are things I forget from time to time (S7).	Excitement: Sometimes they wanted to convey that I was excited or scared here; when my teacher called me up to the board to solve a problem, my hands would shake (S3).	Difficult: When taking exams, difficult questions scare us, especially if there are things we don't know, I get very nervous (S4).	Family Support: My family supports me, so I'm not worried (S8).
Self-Confidence: I trust myself; I usually get good grades, so I don't think I'll get a bad grade on my exam (S7)	Obsessive Thought: <i>I think I will be more successful if I play with my pen, so I constantly play with my pen</i> (S6).	Restlessness: <i>Sometimes I get bored during exams or classes and start playing with my paper and pen (S1).</i> Feeling of Failure: I also feel afraid when I experience feelings of failure (S9).		

*Students have been assigned codes such as S1, S2.

Note: Inappropriate codes are indicated in bold and italics.

Appendix 4

Category and Code Table for Themes Related to Items in the Expanded-Alternative Form

Person-Related Characteristics			Features of the Exam	Family Characteristics
<i>Personal Characteristics</i>	<i>Cognitive Characteristics</i>	<i>Sensory Characteristics</i>	<i>The Difficulty of the Exam</i>	<i>The Family's Attitude</i>
Cheating: During exams, I only look at my own paper, otherwise my teacher will think I'm cheating (S1*).	<i>Note: If I play with my pen during the exam, my concentration is disrupted, so I never play with it (S2).</i>	Test Anxiety: During exams, my hands shake a lot in the first few minutes, then it goes away. I think they shake because I'm scared (S5).	Easy: If the exam is easy, I try to finish it quickly. If the exam is difficult, I can't finish it quickly even if I want to, but it usually seems easy to me (S2).	Family's Reaction: I get anxious during exams because they forbid me from playing with my tablet when I get bad grades (S8).
Time Management: I try to finish the exam quickly and use the remaining time to review the questions and my answers (S8).	Trying to Recall Information: When I get stuck on a question because I can't remember the answers, I start fidgeting (S9).	Excitement: During exams, I usually feel my ears getting warm; they get warm when I'm scared (S10).	Difficult: I wonder if my answers are correct when the exam is difficult because when it's easy, I can guess whether they are correct or not (S7).	Family Support: I think about what my family will say during the exam, but they'll be happy if I get good grades (S6).
Self-confidence: I'm not afraid of doing something wrong on the exam because I don't make mistakes (S4).		<i>Restlessness: In some exams, I get bored and want to finish quickly and go out into the garden (S4).</i>		
		Curiosity: During exams, I sometimes look around because I'm curious about what's going on. When my teacher warns a friend, I turn and look (S1).		

*Students have been assigned codes such as S1, S2.

Note: Inappropriate codes are indicated in bold and italics.

Ekler**Ek 1**

Genişletilmiş-Alternatif ve Klasik Likert Formları için Cronbach Alfa Güvenilirlik Katsayıları

	Tüm Ölçek	Düşünceler	Davranış	Özerk Tepkiler
Genişletilmiş-Alternatif Form	.87	.81	.61	.78
Klasik Form	.88	.83	.64	.76

**p < 0.01

Ek 2

Çocuk ve Ergen Gruplarına Göre Farklı Gruplardan Elde Edilen Cronbach Alfa İç Tutarlılık Katsayılarının Karşılaştırılması

Form	Grup	Tüm Ölçek	Düşünceler	Görev Dışı Davranış	Özerk Tepkiler
Genişletilmiş-Alternatif Form	Çocuk Grubu	.86	.78	.61	.75
	Ergen Grubu	.91	.87	.70	.84
Klasik Form	Child Group	.89	.82	.68	.76
	Adolescent	.84	.83	.55	.75

Ek 3

Klasik Formda Yer Alan Maddelere İlişkin Temalara ait Ait Kategori ve Kod Tablosu

Kişiye Ait Özellikler			Sınava Ait Özellikler	Aileye Ait Özellikler
<i>Kişisel Özellikler</i>	<i>Bilişsel Özellikler</i>	<i>Duyuşsal Özellikler</i>	<i>Sınavın Güçlüğü</i>	<i>Ailenin Tutumu</i>
Kopya Çekmek: Öğretmenim, burada kopya çekenler hani sürekli arkadaşakilere ya da önündeki öğrencilere bakıyor ya o sorulmuş (Ö1*).	Dikkat: Evet dikkatim çok abuk dağıldığından, başka şeyler hakkında düşünerek dalyorum. (Ö8).	Sınav Korkusu: Başarısız olursam annem çok kızar, bence sınavdan korkup korkmadığımız soruluyor (Ö2).	Kolay: Sınavın zor mu kolay mı olacağıyla ilgili düşüncemizi soruyor. Sınav kolaysa beni endişelendiriyor (Ö1).	Ailenin Tepkisi: Benim annem kötü sonuç alırsam bazen kızabiliyor ama az kızıyor. Bence burada ailemizin kızıp kızmayacağını öğrenmek istiyor (Ö5).
Zaman Yönetimi: Sınavda çok fazla hareket etmem zaman kaybı olarak görüyorum, sadece kâğıdıma dönük işlem yaparım (Ö10)	Unutkanlık: Bazen böyle düşünürüm çünkü arada unuttuğum konular oluyor (Ö7).	Heyecan: Burada heyecanlandığım ya da korktuğumuz anlatılmak istenmiş bazen, öğretmenim problem çözmek için tahtaya kaldırdığında ellerim titriyor (Ö3).	Zor: Sınav olurken soruların zor olması bizi korkutur, özellikle bilmediğimiz şeyler varsa çok tedirgin oluyorum (Ö4).	Ailenin Desteği: Ailem beni destekliyor bu nedenle endişelenmiyorum (Ö8).
Özgüven: Kendime güvenirim genellikle iyi notlar alırım bu nedenle sınav sonucumda kötü bir not alacağımı düşünmem (Ö7).	Obsesif Düşünce: Kalemle oynadığımda daha başarılı olacağımı düşünüyorum bu nedenle sürekli kalemimle oynarım (Ö6).	Sıkılganlık: Bazen bende sınavda ya da derste sıkılıp kâğıt- kalemle oynuyorum (Ö1).		
		Başarısızlık Duygusu: Başarısızlık duygusunu yaşadığım zamanlar da korkuyorum (Ö9).		

*Öğrencilere Ö1, Ö2 şeklinde kod verilmiştir.

Not: Amaca uygun olmayan kodlar bold ve italik olarak belirtilmiştir.

Ek 4**Geniřletilmiş-Alternatif Formda Yer Alan Maddelere İliřkin Temalara ait Ait Kategori ve Kod Tablosu**

Kişiyeye Ait Özellikler	Sınav Özellikler	Ait Özellikler	Aileye Ait Özellikler	
Kişisel Özellikler	Bilişsel Özellikler	Duyuşsal Özellikler	Sınavın Güçlüğü	Ailenin Tutumu
Kopya Çekmek: Sınavda sadece kendi kâğıdına bakarım yoksa öğretmenim kopya çektiğimi düşünür (Ö1*). Zaman Yönetimi: Sınavı hızlı bitirmeye çalışırım kalan zamanımda tekrar soruları ve cevaplarımı kontrol ederim (Ö8). Özgüven: Sınavda yanlış bir şey yapmaktan korkmam çünkü hata yapmıyorum (Ö4).	Dikkat: Sınavda kalemimle oynarsam dikkatim dağılır bu nedenle hiçbir zaman oynamam (Ö2). Bilgiyi Getirmeye Çalışma: Cevapları hatırlamadığımda soruda takıldığımda sırada hareket etmeye başlıyorum (Ö9).	Sınav Korkusu: sınavdayken ilk dakikalar ellerim çok titriyor, sonra geçiyor sanırım korktuğum için titriyor (Ö5). Heyecan: Sınavdayken kulaklarımda ısındığını hissediyorum genelde korktuğum zamanlar ısınırlar (Ö10). Sıkılganlık: Bazı sınavlarda sıkıldığım için hızlıca bitirip bahçeye çıkmak istiyorum (Ö4). Merak: Sınavdayken etrafıma bazen Bakıyorum çünkü neler olduğunu merak ediyorum, öğretmenim bir arkadaşımı uyarır dönüp bakıyorum (Ö1).	Kolay: Sınav Kolaysa hızlı bitirmeye çalışırım sınav zorsa istesem de hızlı bitiremem ama genellikle bana kolay geliyor (Ö2). Zor: Cevaplarımda doğru olup Olmadığını sınav zorsa merak ediyorum çünkü kolay olduğunda doğru olup olmadığını tahmin edebiliyorum (Ö7). Ailenin Tepkisi: Sınav olurken endişeleniyorum çünkü kötü not aldığımda tabletle oynamamı yasaklıyorlar (Ö8). Ailenin Desteği: Sınavdayken ailemin ne diyeceğini düşünüyorum ama onlar mutlu olsunlar diye yani iyi not alınca seviniyorlar (Ö6).	

*Öğrencilere Ö1, Ö2 şeklinde kod verilmiştir.

Not: Amaca uygun olmayan kodlar bold ve italik olarak belirtilmiştir.