Article

Not peer-reviewed version

A Triadic Model Integrating Cognitive, Psychomotor, and Affective Dimensions in Adolescent Musicians: The Music Aptitude Self-Efficacy Scale (MASES)

Gülce Coşkun Şentürk , Aslı Kaya * , Çağrı Başbuğ

Posted Date: 2 March 2026

doi: 10.20944/preprints202603.0039.v1

Keywords: music education; music aptitude; adolescence; self-efficacy; scale development; psychometric properties

Article

A Triadic Model Integrating Cognitive, Psychomotor, and Affective Dimensions in Adolescent Musicians: The Music Aptitude Self-Efficacy Scale (MASES)

Gülce Coşkun Şentürk, Aslı Kaya * and Çağrı Başbuğ

Department of Music Education, Faculty of Education, Mugla Sitki Kocman University, Mugla, Türkiye

* Correspondence: aslikaya@mu.edu.tr

Abstract

Due to the highly anxiety-inducing nature of music aptitude tests or exams, there is a need for a specialized tool to measure adolescents' beliefs about their musical potential. This study aims to develop and validate the Music Aptitude Self-Efficacy Scale (MASES) for adolescent music teacher candidates preparing for aptitude tests. The study group consisted of 383 students (ages 16–18) preparing for the Music Education Department entrance exams at the Education Faculties in Türkiye. Split-sample method was used for Exploratory Factor Analysis (n=199) and Confirmatory Factor Analysis (n=184). The EFA results have revealed a three-factor structure: Cognitive-Auditory Competence, Psychomotor- Performance Competence, and Affective Regulation Competence. CFA results, this three-factor model has excellent fit indices ($\chi 2$/df = 1.24, RMSEA =. 036, CFI =. has been validated with 976). The scale has high internal consistency and composite reliability (Cronbach α =. 87, McDonald's ω =. 92). The criterion-related validity analysis revealed significant positive correlations (r = .586) between the MASES and the Questionnaire for Measuring Self-Efficacy in Youths. The findings indicate that MASES is a valid and reliable tool that can be used by music educators to identify student's self-efficacy levels and to design guidance programs for aptitude tests or exams.

Keywords: music education; music aptitude; adolescence; self-efficacy; scale development; psychometric properties

1. Introduction

Music is not just entertainment; it is a complex auditory stimulus that activates various cognitive processing systems (Sust et al., 2026). Music education is a multidimensional and dynamic process requiring the simultaneous development of an individual's cognitive, affective, and psychomotor skills (Hallam, 2010). Identifying particularly talented individuals and guiding them toward professional music education is one of the most critical agenda items for music educators (Strauss et al., 2024; Tunca et al., 2024). Although talent is considered an innate potential, transforming this potential into performance is directly related to the individual's motivation, anxiety management, and, most importantly, their belief in their capacity (McPherson & McCormick, 2006).

1.1. The Nature and Psychological Dimension of Aptitude Tests

Aptitude test scores are generally interpreted similarly for candidates with the same overall score. However, research has found evidence of differences among candidates in strategies and the consistent application of proper procedures during testing, and it is clear that each student, despite having different strategies and skill levels, is taking those exams for the same purpose (Hauenstein et al., 2024; Hambrick et al., 2023; Christiner et al., 2023; Sideridis,& Alahmadi, 2022). Additionally, talent exams, the most important threshold for transitioning to professional music education, are challenging processes that test not only candidates' musical auditory and performance skills but also

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2026 doi:10.20944/preprints202603.0039.v1

their capacity to perform under high stress. Although there are numerous studies in the literature on musical performance anxiety (Kenny, 2011) and general academic self-efficacy, the limited number of measurement tools specifically focusing on the "aptitude test process" is noteworthy. However, in a candidate's success in the aptitude test, their self-efficacy belief in their aptitude is as influential as the aptitude itself.

1.2. The Türkiye Context: Music Aptitude Tests for Transition to Higher Education

In Türkiye, admission to Music Education Departments within Faculties of Education, which provide professional music education and form the foundation of the music teacher training process, is generally carried out through multi-stage "Special Talent Exams" (auditory perception, sight-reading, instrumental performance, singing, etc.) organized by universities, attended by high school graduates (16-18 years old) and young adults who have graduated from high school. These highly competitive entrance exams for higher education expect candidates to simultaneously demonstrate quick reflexes, technical proficiency, and theoretical knowledge in front of an expert jury. However, the current examination evaluation system focuses solely on the candidate's musical ability, while excluding " aptitude test self-efficacy," which determines the capacity to demonstrate this ability under intense pressure. This situation can lead to candidates, despite possessing high musical potential and technical skills, being excluded from the system due to a decline in their performance in front of the jury caused by low self-efficacy.

1.3. Conceptual Model: The Multidimensional Structure of Music Aptitude Self-Efficacy

In this study, self-efficacy perception regarding music aptitude tests was modeled not as a one-dimensional structure, but as a dynamic interaction of cognitive, psychomotor, and affective processes. The three-factor structure obtained from Exploratory Factor Analysis (EFA) was conceptualized within the framework of the Music Aptitude Self-Efficacy Triadic Model, which aligns with fundamental theories in educational psychology (See Figure 1).

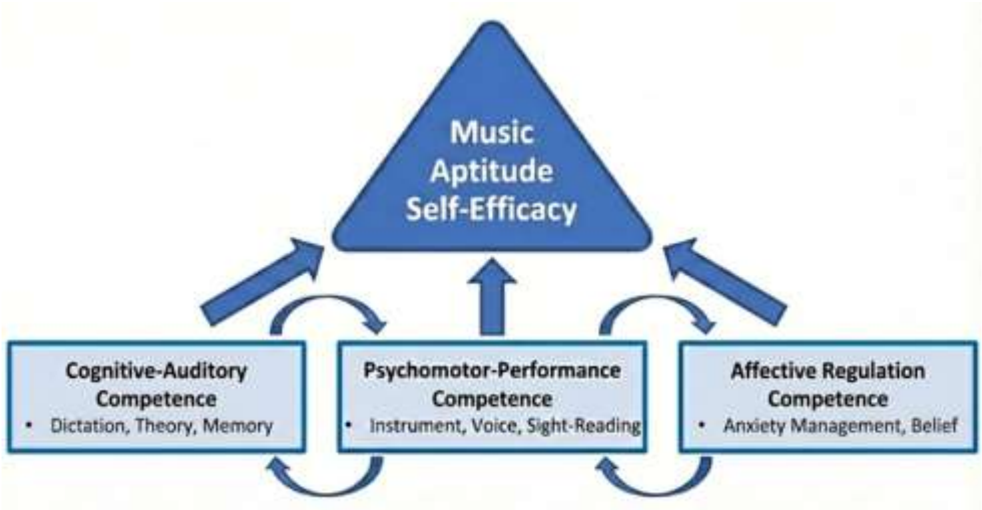


Figure 1. Music Aptitude Self-Efficacy Triadic Model.

The model is constructed from a synthesis of Cognitive (Gordon/Audiation), Psychomotor (Simpson/Execution), and Affective (Bandura/Anxiety Management) theories. The arrows represent the dynamic interaction and reciprocal reinforcement between the dimensions. This model is based on a synthesis of Bandura's (1986) Social Cognitive Theory and Bloom's (1956) taxonomy of educational goals (Cognitive, Affective, Psychomotor) in the context of music education. The components of the model are detailed below.

1.3.1. Cognitive-Auditory Efficacy:

The first dimension of the model encompasses the processes of mentally processing, encoding, and memorizing musical data. This dimension is based on Gordon's (2012) "Audiation" theory. The candidate's ability to write melodic dictation, recognize intervals, and master music theory is related to Bandura's "Cognitive Readiness" as a source of self-efficacy (Zimmerman, 2000). This high level of self-efficacy reflects the presence of mental schemas necessary for the candidate to manage uncertainty during the aptitude exam (McPherson & McCormick, 2006).

1.3.2. Psychomotor-Performance Efficacy:

The second dimension involves the transformation of musical thought into physical action, such as vocal or instrumental performance. This dimension is related to the complex overt response level in Simpson's (1972) psychomotor domain taxonomy. The candidate's mastery of the instrument, belief in overcoming technical difficulties (seeing, etc.), and bodily control are based on Bandura's (1997) "Experiences of Mastery". The belief in one's technical competence in musical performance is the strongest predictor of minimizing the fear of making mistakes during an exam (Hallam, 2010).

1.3.3. Affective Regulation & Readiness:

The third dimension encompasses the candidate's ability to manage test anxiety, motivational orientation, and perception of social support. This dimension is directly related to Bandura's (1997) principle of "Physiological and Affective States" influencing self-efficacy. The ability to cope with performance anxiety and the belief in one's ability represent the "valuing" and "organizing" levels of the affective domain as defined by Krathwohl et al. (1964). Ritchie and Williamon (2012) emphasize that high affective self-efficacy increases resilience in musical performance.

1.4. Gaps in the Literature and the Importance of Research

A review of the literature on music education and performance reveals that the decisive role of self-efficacy and motivational processes in performance success has been strongly demonstrated by McPherson and McCormick (2006) and Ritchie and Williamon (2011). However, these studies, which focus on students currently receiving conservatory training, are limited in their ability to measure the "high- stakes" exam psychology experienced by candidates at the entry stage of the system, and the specific cognitive requirements of these exams, such as auditory (melodic dictation/theory) and instantaneous deciphering, using a holistic model. Similarly, the measurement tools developed by Zelenak (2015) focus more on determining general class participation and motivational orientations, while not offering a diagnostic framework for predicting the outcome of talent exams. On the other hand, the most widely used Kenny Music Performance Anxiety Inventory (K-MPAI) (Kenny, 2011) addresses the performance process from a negative perspective, viewing "anxiety" as a pathological condition. However, current positive psychology approaches argue that success is explained not only by the "absence of anxiety" but also by the "presence of belief in one's ability (self-efficacy)." In this context, the MASES (Music Aptitude Self-Efficacy Scale) fills a significant methodological and theoretical gap in the literature as it is the first measurement tool to address the exam process not solely in terms of performance or anxiety, but within the framework of the Triadic Model, built upon Cognitive, Psychomotor, and Affective pillars.

2. Materials and Methods

2.1. Research Design

This research is a methodological study aiming to develop a valid and reliable measurement tool to assess the self-efficacy perceptions of individuals preparing for music aptitude test or exams. In the scale development process, an Exploratory Sequential Design (ESD) was adopted, combining qualitative and quantitative methods (Creswell & Plano Clark, 2007). In the first stage, a literature review was conducted, expert opinions were obtained using the Delphi Technique, and an item pool was created through cognitive interviews. In the second stage, psychometric analyses (EFA, CFA, Reliability, Validity) were performed on the obtained data.

2.2. Participants

The study population consists of 11th and 12th grade students studying in the music departments of Fine Arts High Schools who aim to take the entrance exam for the music education departments of universities in Türkiye. Purposive Sampling method was preferred in accordance with the principles of accessibility and suitability for purpose. Data was collected from a total of 406 participants, but after data cleaning (see results section), the analyses were conducted on 383 participants. To test the construct validity of the scale and to ensure the cross-validity of the model, the “Split-Sample Method” suggested in the literature was applied (Anderson & Gerbing, 1988). Accordingly, the dataset was randomly divided into two subgroups:

- Grup 1 (EFA group): This group (n_1), used for exploratory factor analysis, consisted of 199 participants (Mean Age = 17.11; 63.8% Female, 36.2% Male; 45.7% 11th grade, 54.3% 12th grade).
- Grup 2 (CFA group): This group (n_2), used for confirmatory factor analysis, consisted of 184 participants (Mean Age = 17.09; 61.4% Female, 38.6% Male; 46.7% 11th grade, 53.3% 12th grade).
- Grup 3 (Validity group): This group (N), used for validity analysis, consisted of 380 participants (Mean age = 17.10; 62.6% female, 37.4% male; 45.8% 11th grade, 54.2% 12th grade).

The sample size in the study meets the rule suggested by Comrey and Lee (2013) and Tabachnick and Fidell (2013), which states that the sample size in both subgroups should be at least 5 times the number of items (k=35) ($N/k > 5$), and is therefore considered “sufficient” for factor analysis.

2.3. Procedure

To create the item pool of the scale and ensure its content validity, the Delphi Technique, which is accepted in the literature as a systematic consensus method, was used (Linstone & Turoff, 1975). This process allowed experts to express their opinions independently and anonymously, thus avoiding group effect (halo effect) (Dalkey & Helmer, 1963). The scale development process was carried out by following the standard procedures suggested by DeVellis (2017):

2.3.1. Item Generation

The following processes were carried out during the item pooling phase:

- Participants and Sampling: To create the item pool for the scale and to test the cultural suitability of the theoretical structure (Triple Model), “Criterion Sampling” from “Purposive Sampling” methods was used (Patton, 2014). The study group consisted of 10 participants who were preparing for or had experienced the entrance exams (aptitude exams) for higher education in the field of music. To ensure a heterogeneous group structure, participants were selected from among graduates who had completed their exam experience (n=3), 12th-grade students in their exam year (n=4), and 11th-grade students in the preparation process (n=3) (Table 1).

Table 1. The profile of study group (Qualitative Phase).

Code	Gender	Grade	Instrument	Musical Background (year)	Exam Experience
P1	F	11.	Piano	4	Preparation
P2	M	12.	Violin	6	Preparation
P3	F	Graduated	Baglama	3	3 times
P4	M	12.	Guitar	4	Preparation
P5	M	12.	Flute	4	Preparation
P6	M	11.	Piano	3	Preparation

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2026 doi:10.20944/preprints202603.0039.v1

P7	F	12.	Viola	6	Preparation
P8	F	11.	Guitar	5	Preparation
P9	M	Graduated	Cello	6	2 times
P10	F	Graduated	Violin	8	4 times

- Data Collection Process: In data collection, the "Semi-Structured Focus Group Interview" technique was preferred, allowing participants to interact with each other and generate in-depth data (Krueger & Casey, 2015). Since the number of participants (N=10) exceeded the ideal group size suggested in the literature and could reduce interaction, the interviews were conducted in two separate sessions of 5 people each (Morgan, 1997). Each session was conducted outside of class time in a quiet and distraction-free environment in the research school building after obtaining parental consent, and lasted an average of 50-60 minutes. During the interviews, the researcher acted as the moderator, while an assistant researcher observed and noted nonverbal responses during the process.
- Recording and Transcription: To prevent data loss and ensure the reliability of the analysis, all interviews were recorded using an audio recorder after obtaining "Informed Consent Forms" from the participants. During the interviews, the researcher also handwritten field notes containing keywords and key points (Creswell, 2013). Following the interviews, the audio recordings were transcribed verbatim without any corrections, resulting in a raw dataset of 24 pages.
- Interview Protocol: In the interviews, a "Semi-Structured Interview Form" based on the theories of Bandura, Gordon, and Simpson was used. The questions were structured from general to specific to reveal the candidates' perceptions of the cognitive, psychomotor, and affective dimensions of the aptitude test. In addition, probe questions were used to deepen the participants' responses (Yıldırım & Şimşek, 2018). The main questions and probe questions asked were as follows:

1. Cognitive-Auditory Dimension (Focused Gordon):

- Main question: "When considering the auditory stage of aptitude tests, what are the moments that challenge you the most or make you think, 'I can't do it'?"
- Probe question: "What kind of confusion do you experience in your mind, especially when dictating or hearing intervals?"

2. Psychomotor-Performance Dimension (Focused Simpson):

- Main question: " "When playing your instrument or singing, what are your thoughts on technical difficulties that might affect your performance in the aptitude test environment?"
- Probe question: "What do you feel in your body when asked to sight-read a piece you don't know?"

3. Affective Regulation Dimension (Focused Bandura):

- Main question: "How would you describe the moment just before entering the exam room and the first moment you stand before the jury?"
- Probe question: "What do you tell yourself to manage your nervousness? How does hearing the other candidates affect you?"
- • Data Analysis: The descriptive analysis method was applied to the obtained data. The data were coded according to predetermined theoretical themes (cognitive, psychomotor, and affective); inductive coding was used for new situations that did not fit the theoretical framework but were frequently emphasized by the participants (Miles, Huberman, & Saldaña, 2014). The item generation matrix is presented in Table 2.

 doi:10.20944/preprints202603.0039.v1

Table 2. Item generation matrix.

Student Statement (Direct Quote)[1]	Emerging Theme	Related Theory	Generated Draft Item
"…, I cannot discern the high pitches notes in between." (P2)	Auditory Discrimination	**Gordon** (Harmonic Audiation)	*Item 18: I am able to discriminate between the pitches during the polyphonic auditory test.*
"… I cannot sustain the song until the end." (P9)	Physiological Control	**Simpson** (Mechanism)	*Item 28: I am able to maintain technically correct breath and vocal control, even under the stress of aptitude examinations.*
"… thinking [this], I feel discouraged." (P1)	Social Comparison	**Bandura** (Vicarious Experience)	*Item 5: Hearing the performances of other candidates does not negatively affect my belief in my own success.*

[1] Details are given in Table S1.

As a result, a broad draft pool of 47 items was created by combining theoretical points from the literature and statements derived from student opinions. The generated items are listed in Table S1.

2.3.2. Content Validity and the Delphi Technique

The draft form was presented to an expert panel of five expert in the fields of music education, measurement and evaluation, and psychological counseling. The expert panel was formed from academics and experts with an average of 15 years of professional experience, selected from the fields of Music Education (n=2), Educational Psychology (n=1), Measurement and Evaluation (n=1), and Examiner/Teacher (n=1), to ensure multidimensional content validity (Appendix A). Lynn (1986) states that a minimum of 3 experts are required for content validity studies and that 5 experts is a sufficient and reliable number for establishing content validity. Experts were asked to evaluate the items using the Davis Technique (1: Not Relevant - 4: Highly Relevant) (Davis, 1992). As a result of the evaluation, the Content Validity Index (S-CVI) was calculated as .86, which is above the threshold of .80 (Polit & Beck, 2006), thus concluding that the items have achieved content validity. In line with expert recommendations, 12 items were removed from the pool, and some technical terms were simplified (Table S2).

2.3.3. Face Validity and Pilot Study

To test the comprehensibility of the 35-item draft form, whose content validity was ensured based on expert opinion, a cognitive interviewing study was conducted among the target candidate (16-18 age group). In this stage, the "Think Aloud" and "Verbal Probe" techniques suggested by Willis (2005) were used.

The study group consisted of 15 students selected from different grade levels (11th and 12th grade) and different instrument groups. In one-on-one interviews lasting approximately 20 minutes each, students were asked to read the items aloud, and their meaning-making processes were explored through questions such as, "What does this item bring to mind?" and "How would you express this word in your words?".

As a result of the interviews, technical terms found in academic literature but potentially ambiguous for the target audience's age level were identified. For instance, it was determined that students perceived the expression 'playing clearly and distinctly' more accurately than the term 'articulation,' and 'the tone/mode of the melody' more accurately than 'tonality'. Five items that caused semantic confusion were rephrased, technical terms were simplified, and the scale's instructions were adapted to suit students' music aptitude exam psychology, making it ready for the pilot application (Table S3). The 35 statements included in the draft scale are designed to be rated on a scale of 1 to 5 (1: Never, 5: Completely) to best reflect the level of competence in that situation.

2.4. Data Analysis

In the statistical analysis of the data, JASP 0.95.4 was used for descriptive statistics and Exploratory Factor Analysis (EFA), and SPSS 26.0 and AMOS 24.0 (and the JASP CFA module) were used in an integrated manner for Confirmatory Factor Analysis (CFA) and reliability calculations. The analysis process was carried out following a four-stage protocol:

2.4.1. Data Screening and Assumption Testing:

Prior to analysis, the dataset was scanned for erroneous entries and missing data. The multivariate normality assumption and outliers were examined using the Mahalanobis distance criterion (p < .001); outliers exceeding the critical threshold (χ^2) were removed from the dataset (Tabachnick & Fidell, 2013). The normal distribution of the data was checked using the criteria of skewness and kurtosis coefficients being within the ±1.5 range (George & Mallery, 2019). In addition, sample adequacy, a prerequisite for factor analysis, was tested with the Kaiser-Meyer-Olkin (KMO) test (>.80), and the suitability of the correlation matrix between variables was tested with the Bartlett's Test of Sphericity ($p < .001$).

2.4.2. Structural Exploration: Exploratory Factor Analysis (EFA)

The first sub-sample, consisting of 199 participants, was utilized for the Exploratory Factor Analysis (EFA). To reveal the latent structure of the scale, Principal Axis Factoring (PAF)—which focuses on common variance and is considered more robust in identifying theoretical structures—was preferred over Principal Component Analysis (PCA) (Costello & Osborne, 2005). Since the cognitive, psychomotor, and affective dimensions comprising music talent were theoretically expected to be interrelated, Promax (Kappa=4), an oblique rotation method, was employed instead of orthogonal rotation. Rather than relying on a single criterion to determine the number of factors, (a) the Eigenvalue > 1 rule, (b) Parallel Analysis results, and (c) the Scree Plot were evaluated collectively. As criteria for item deletion, factor loadings below .40 or a loading difference of less than .10 between two factors (cross-loading) were established (Hair et al., 2019)."

2.4.3. Structural Validation: Confirmatory Factor Analysis (CFA):

The second sub-sample, consisting of 184 participants, was utilized for the Confirmatory Factor Analysis (CFA). In the literature, the number of participants per observed variable or item (N:p ratio) is considered a critical criterion in determining sample size. Bentler and Chou (1987) state that at least 5 participants per item (5:1 ratio) is sufficient for reliable parameter estimations. In this study, this ratio was calculated as approximately 8.8:1 (184/21) for the 21-item final form. Furthermore, Tabachnick and Fidell (2013) and Hair et al. (2019) state that sample sizes between 150 and 200 are within acceptable limits for CFA in simple models with strong factor loadings. Accordingly, the sample size was evaluated to have sufficient statistical power for model validation.

To test the fit of the three-factor structure (cognitive, psychomotor, and affective) identified through the EFA, a measurement model was established. The measurement model is a structural equation model that defines the relationship of each latent variable (factor) with the observed variables (items) and accounts for measurement errors (Byrne, 2010). In the parameter estimation of the model, the Maximum Likelihood (ML) estimation method was preferred, as the data substantially met the assumption of multivariate normality. For model specification, a first-order CFA was applied, in which the covariances between factors were freely estimated and the variance of the first item of each latent variable was fixed to 1 (unit loading identification).

To evaluate the model's fit to the dataset, it is recommended to report multiple indices reflecting different fit characteristics rather than a single index (Kline, 2015; Brown, 2015). Accordingly, absolute goodness-of-fit indices such as χ^2/df (Chi-square/degrees of freedom), RMSEA (Root Mean Square Error of Approximation), and SRMR (Standardized Root Mean Square Residual) were examined, and comparative goodness-of-fit indices such as CFI (Comparative Fit Index) and TLI (Tucker-Lewis

Index) were investigated. In interpreting the goodness-of-fit values, the combinational rule strategy proposed by Hu and Bentler (1999) was adopted. The accepted fit indices reference range values (Browne et al., 1993; Marsh et al., 2004) are presented in Table 3.

Table 3. Fit indices.

Fit indices	Excellent fit	Acceptable fit	References
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 \leq \chi^2/df \leq 4$	Kline (2015)
RMSEA	≤ .06	≤ .08	Hu & Bentler (1999)
SRMR	≤ .05	≤ .08	Byrne (2010)
CFI	≥ .95	≥ .90	Hu & Bentler (1999)
TLI	≥ .95	≥ .90	Bentler & Bonett (1980)

2.4.4. Evidence for Reliability and Validity:

In the reliability analysis of the scale, in addition to the traditional Cronbach's Alpha (α) coefficient, McDonald's Omega (ω) and Composite Reliability (CR) coefficients—which provide more robust results as they are based on factor loadings—were reported. For convergent validity, the criteria established were an Average Variance Extracted (AVE) value above .50, a Composite Reliability (CR) value above .70, and the fulfillment of the CR > AVE condition (Fornell & Larcker, 1981). Discriminant validity was examined based on the criterion that the square root of the AVE values should be higher than the inter-factor correlations.

2.5. Concurrent Validity

To test the concurrent validity of the developed the MASES, Self-Efficacy Questionnaire for Children (SEQ-C), developed by Muris (2001) was utilized. The SEQ-C adapted into Turkish by Çelikkaleli et al. (2006) as the Questionnaire for Measuring Self-Efficacy in Youths. The SEQ-C is a 5-point Likert-type (1: Not at all, 5: Very well) instrument consisting of 23 items. The scale measures adolescents' general competence beliefs and consists of three sub-dimensions: Academic Self-Efficacy (8 items), Social Self-Efficacy (8 items), and Emotional Self-Efficacy (7 items). In the original and adaptation studies, internal consistency coefficients were reported to range between .64 and .71. Scores on the scale range from a minimum of 23 to a maximum of 115. All items in the SEQ-C are scored positively, where higher scores indicate a higher level of self-efficacy belief.

Although the construct validity analyses of the scale were conducted using split samples, criterion-related validity analyses, which test the relationship of the scale with external criteria, were performed on the entire dataset (N = 383) to increase statistical power and ensure the stability of parameter estimations (Cohen, 1988; Tabachnick & Fidell, 2013). And, pearson correlation coefficients were examined between the total MASES scores, calculated from the 21 items in the final form, and the scores on the SEQ-C.

3. Results

3.1. Data Preparation and Pre-Analysisve

Before the analysis process, the raw dataset obtained from 406 participants was rigorously examined for data cleaning, outliers, and normality assumptions. In the first stage, Mahalanobis distance was calculated to identify multivariate outliers ($p < .001$), and 23 participants exceeding the critical value (χ^2 = 66,62) were excluded from the analysis (Tabachnick & Fidell, 2013).

Normality tests conducted on the remaining dataset (N = 383) showed that the items' skewness (-1.008 to .710) and kurtosis (-1.096 to .991) values were within acceptable limits (±1.5). However, Item 5 (I believe that those around me think I will succeed in the aptitude test) was removed from the dataset at this stage due to its kurtosis value being excessively high (1.826), which distorted the distribution, and its limited contribution to the theoretical structure (George & Mallery, 2019).

Before the factor analysis, item analysis was conducted to test the discrimination levels of the items. The Critical Ratio (CR) method confirmed strong item discrimination, showing significantly high t-values for all items (ranging from 3.94 to 13.71; $p < .001$) (Ebel & Frisbie, 1991). When checking the Corrected Item-Total Correlation (CITC), 9 items (i2, i8, i12, i16, i19, i26, i27, i28, i29; r range: .160 - .389) that did not meet the discrimination coefficient criterion of .40 were eliminated (Nunnally & Bernstein, 1994). Following this elimination, it was determined that the discrimination coefficients of the remaining 25 items ranged from .416 to .644, and the Cronbach's Alpha internal consistency coefficient of the scale increased from .899 to .923 (Table S4).

To test construct validity and improve the generalizability of the model, the two-step approach suggested in the literature was adopted (Anderson & Gerbing, 1988; Worthington & Whittaker, 2006; DeVellis, 2017). In this context, the dataset was randomly divided into two subsamples; the first subsample ($n_1 = 199$) was used for Exploratory Factor Analysis (EFA) to reveal the factor structure of the scale, and the second subsample ($n_2 = 184$) was used for Confirmatory Factor Analysis (CFA) to confirm the revealed structure.

3.2. Exploratory Factor Analysis (EFA)

To determine the factor structure of the scale, which was reduced to 25 items, EFA was applied to the first subsample ($n_1 = 199$) using the JASP 0.95.4 program. Principal Axis Factoring (PAF) was preferred as the inference method due to its effectiveness in revealing latent variables, and Promax (Oblique) was chosen as the rotation method because the factors are theoretically expected to be related (cognitive, psychomotor, and affective processes) (Costello & Osborne, 2005; Hair et al., 2019).

An iterative approach was followed in the analysis process. In the initial analysis, the parallel analysis results pointed to a three-factor structure; however, upon examining the factor loadings, items with low loading values (< .40) were identified. Following the methodology of Hair et al. (2019), four items with low loadings (i1, i15, i20, i30) were removed from the analysis, and no cross-loading items were identified during this process.

The final EFA results obtained with 21 items showed that the data were perfectly suitable for factor analysis (KMO = .937; Bartlett's Test $\chi^2 = 4220.656$, df = 210, $p < .001$). Parallel analysis and the Scree Plot confirmed the three-factor structure, consistent with theoretical expectations (See Figure 2). This three-factor structure explains 75.4% of the total variance.

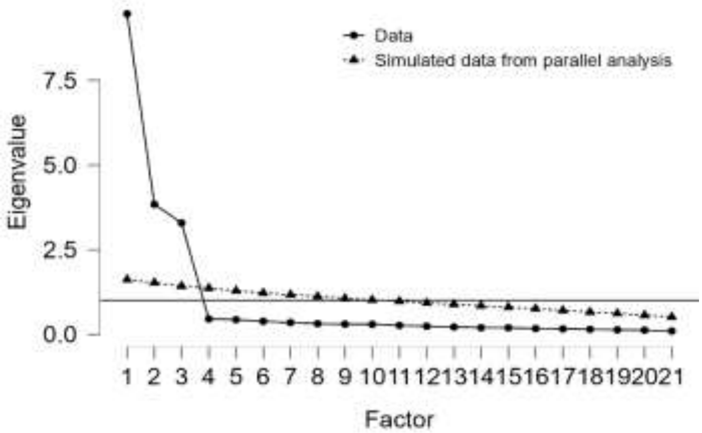


Figure 2. Scree plot of EFA.

When the model's fit indices were examined, CFI = .975, TLI = .964, and SRMR = .018 indicated good fit, while RMSEA = .058 was within acceptable limits for the EFA stage (Hu & Bentler, 1999). The factor loadings, common variances, and dimension assignments are shown Table 4.

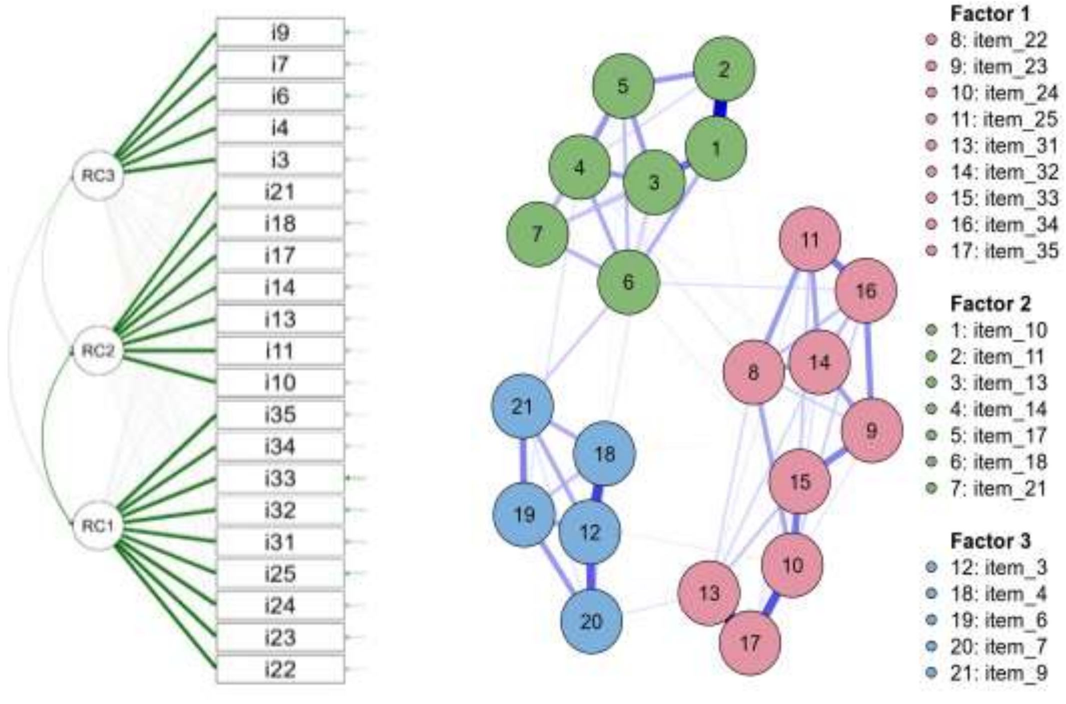
 doi:10.20944/preprints202603.0039.v1

Table 4. Factor loadings and communalities from EFA of the scale 21-Item (N = 199).

Items	Factor 1	Factor 2	Factor 3	h^2
i24	.918			.819
i32	.900			.817
i35	.899			.793
i23	.896			.793
i33	.896			.772
i34	.873			.789
i22	.872			.815
i31	.872			.766
i25	.863			.766
i13		.900		.769
i10		.886		.786
i11		.869		.739
i17		.863		.741
i18		.818		.705
i21		.790		.606
i14		.767		.686
i3			.914	.818
i4			.863	.745
i7			.841	.717
i6			.839	.709
i9			.822	.683
Eigenvalues	9.227	3.582	3.025	
% of Variance	43.9	17.1	14.4	
Cumulative %	43.9	61.0	75.4	75,4 (Total)
McDonald's ω	0.971	0.946	0.932	0.927 (Total)
Cronbach's α	0.971	0.945	0.931	0.933 (Total)

*Note.*Applied rotation method is promax. Factor loadings < 0.40 are suppressed; h^2 = communalities. Barlett test (χ^2 = 4220.656; df = 210; p < .001) ve KMO testi (KMO = 0,937).

According to Table 4, the factor loadings of all items are above the threshold value of .40, and ranging from 0.767 to 0.918. All items on the scale showed sufficient common variance (range: 0.606–0.882). The scale resulted in a final 21-item structure with a clear three-factor structure explaining 75.4% of the total variance. Harman's single-factor test revealed that the first unrotated factor explained 42.9% of the variance, indicating that common method bias was not a significant concern in the current study. The resulting structure offers the best balance in terms of statistical fit and theoretical interpretability (Hair et al., 2019; Tabachnick & Fidell, 2019; Stevens, 2002; Comrey & Lee, 2013; Hayton et al., 2004). The factors were theoretically named. Factor 1 (Psychomotor- Performance Competence) consists of 9 items, Factor 2 (Cognitive-Auditory Competence) consists of 7 items, and Factor 3 (Affective Regulation Competence) consists of 5 items. Cronbach's α values are 0.971, 0.945, and 0.931, respectively, with an overall scale α value of 0.933. Path diagram and network graphic of the scale shown Figure 3.



(**a**) (**b**)

Figure 3. (**a**) Path diagram of Triadic Model (MASES); (**b**) Network graphic of Triadic Model (MASES).

As shown in Figure 3, the inter-item correlations within the factors of the three-dimensional, 21-item scale ranged between 0,001 and 0,504, remaining well below the 0,90 threshold for redundancy.

3.3. Confirmatory Factor Analysis (CFA)

Before conducting the Confirmatory Factor Analysis, the data distribution characteristics of the subsample (n_2 = 184) used in the analysis were examined. The condition of multivariate normality, which is the fundamental assumption of the Maximum Likelihood (ML) estimation method, was checked. It was observed that the skewness and kurtosis values of the items ranged within the ±1.5 range (Tabachnick & Fidell, 2013) and Mardia's multivariate normality coefficients were within acceptable limits for the analysis. Furthermore, based on the Mahalanobis distance criterion ($p < .001$), it was determined that there were no multivariate outliers in this subgroup that would bias the analysis.

To test the accuracy of the three-factor structure identified in the EFA phase, Confirmatory Factor Analysis (CFA) was performed on the second subsample (n_2 = 184). Confirmatory Factor Analysis fit indices values shown Table 5.

Table 5. Confirmatory factor analysis fit indices.

Model	χ^2	*df*	χ^2/df	CFI	TLI	SRMR	RMSEA
Triadic	220.602*	186	1.19	0.984	0.982	0.042	0.032**

Note. *p = .042; **(90% *CI* [.007 - .047]).

According to Table 5, the goodness-of-fit values of the model were observed, and the Chi-Square value was statistically significant ($\chi^2 = 220.602$, $df = 186$, $p = .042$). However, since the sensitivity of the Chi-Square value to sample size is known (Kline, 2015), the χ^2/df ratio was used to evaluate the model fit. The calculated ratio of 1.19 ($\chi^2/df < 2$) indicates that the model shows excellent fit. When other fit indices were examined, it was observed that the comparative fit indices *CFI* = .984 and *TLI* = .982 were above the threshold of .95; and the absolute fit indices *RMSEA* was .032 (90% *CI* [.007 - .047]), below the .05 threshold. Furthermore, the *SRMR* value was calculated as .042, which is within the limits of perfect fit (Hu & Bentler, 1999; Browne et al., 1993). Since all primary indices (*CFI, TLI, RMSEA*) indicated excellent fit, it was concluded that the model was validated without the need for modification. These results demonstrate that the three-factor theoretical structure (Triadic Model) of the MASES scale is highly compatible with the data and that construct validity has been established

3.4. Reliability and Validity

Internal consistency, composite reliability, and convergent validity analyses were conducted to determine the psychometric quality of the scale. In the reliability analyses, Cronbach's Alpha (α) coefficient, the McDonald's Omega (ω) coefficient—which provides more robust results as it is based on factor loadings—was examined. For convergent validity, the Average Variance Extracted (AVE) and Composite Reliability (CR) values were investigated. The reliability and validity results of the scale are presented in Table 6."

Table 6. Reliability and convergent validity results.

Factor	k	α	ω	CR	AVE
*F_1	9	.930	.930	.93	.596
*F_2	7	.911	.912	.88	.600
*F_3	5	.863	.863	.86	.561
Total	21	.891	.939		

Note. α: Internal consistency coefficient; ω: Omega reliability coefficient; CR: Composite Reliability; AVE: Average Variance Extracted; F_1: Psychomotor- Performance Competence, F_2: Cognitive-Auditory Competence, F_3: Affective Regulation Competence. *All factor variances: $p < .001$.

As shown in Table 6, the α (.863 - .930) and ω (.863 - .930) values for all factors and the total scale are well above the .70 threshold, indicating a high level of reliability (Hair et al., 2019). In the literature, an AVE value above .50 and a CR value above .70 are expected. The AVE values of factors F1 (.596), F2 (.600), and F3 (.561) directly satisfy this condition. Furthermore, considering that the CR values of all factors are above .70 and the factor loadings are statistically significant ($p < .001$), it was determined that the scale has convergent validity. In conclusion, the findings confirm that the developed scale both provides consistent measurements (Reliability) and accurately represents the construct it aims to measure (Validity).

3.5. Criterion-Related Validity

To test the concurrent validity of the developed MASES, the SEQ-C developed by Muris (2001) and adapted into Turkish by Çelikkaleli et al. (2006) was utilized as a criterion. This scale measures adolescents' general competence beliefs and consists of three sub-dimensions: Academic, Social, and Emotional. The SEQ-C was incorporated into the study as an external criterion to validate the theoretical relationship between musical self-efficacy and general self-efficacy beliefs.

Mahalanobis Distance values were calculated to remove multivariate outliers from the dataset. The critical threshold was determined as 49.73, based on the number of items included in the analysis ($df = 23$) and the significance level ($p < .001$) (Tabachnick & Fidell, 2013). As a result of the review, the dataset of 3 participants exceeding the critical threshold and exhibiting outlier characteristics was removed; the analyses were conducted on the remaining 380 data cases. When the normality distribution of the remaining data was examined, it was observed that the skewness and kurtosis coefficients varied within the ±1.5 range. These values indicate that the data meet the assumption of normal distribution and that parametric tests (Pearson Correlation, etc.) are applicable (George & Mallery, 2019; Kline, 2015). The results of the reliability analysis of the criterion scale are presented in Table 7.

 doi:10.20944/preprints202603.0039.v1

Table 7. Criterion scale reliability analysis results.

Scale	*k*	Cronbach's α	McDonald's ω
Academic Self-Efficacy	8	.830	.831
Social Self-Efficacy	8	.817	.818
Emotional Self-Efficacy	7	.862	.863
the SEQ-C	23	.920	.921

According to Table 7, the Cronbach Alpha (α) and McDonald's Omega (ω) coefficients range from .817 to .921. The fact that the obtained coefficients are well above the .70 confidence threshold accepted in the literature (Nunnally, 1978; Hair et al., 2019) indicates that the criterion scale has a high level of internal consistency within the scope of this study. This finding confirms that the correlation coefficients calculated between MASES and the Adolescent Self-Efficacy Expectation Scale are reliable and interpretable relationships, not resulting from measurement error. The results of the Pearson Correlation Analysis between the MASES and the SEQ-C are presented in Table 8.

Table 8. Pearson correlation analysis results.

Variables	Academic Self-Efficacy	Social Self-Efficacy	Emotional Self-Efficacy	the SEQ-C
F_1	.320**	.309**	.322**	.364**
F_2	.345**	.401**	.397**	.437**
F_3	.334**	.412**	.406**	.442**
the MASES	.463**	.533**	.532**	.586**

Note. N = 380; ***p* < .001. F_1: Psychomotor- Performance Competence, F_2: Cognitive-Auditory Competence, F_3: Affective Regulation Competence.

Table 8 shows that there is a statistically significant, positive, and moderate-to-high correlation between the MASES total score and the SEQ-C total score (r = .586, p < .001). In the literature, correlation values between .30 and .70 are considered ideal, as they indicate that the developed scale theoretically overlaps with the criterion scale while measures a different construct (Cohen, 1988; Büyüköztürk, 2011).

When examined at the sub-dimension level, it was observed that all sub-dimensions of the MASES (F1, F2, F3) exhibited positive and significant correlations with the sub-dimensions of the SEQ-C (Academic, Social, Emotional) ranging from .309 to .533. Notably, the strong relationship between the MASES total score and the Social Self-Efficacy (r = .533) and Emotional Self-Efficacy (r = .532) dimensions demonstrates that musical aptitude self-efficacy is not only a technical skill but is also closely linked to social and emotional self-efficacy beliefs. These coefficients provide strong evidence that criterion-related validity has been established in addition to the construct validity of the MASES. These coefficients provide strong evidence that MASES have both construct validity and criterion-related validity.

4. Discussion

The main objective of this study is to develop a valid and reliable measurement tool that can measure the musical self-efficacy perceptions of adolescents (16-18 years old) preparing for music aptitude tests or exams. The research findings revealed that the developed scale (the MASES) exhibits a Triadic Model structure consisting of Cognitive-Auditory Competence, Psychomotor-Performance Competence, and Affective Regulation Competence, and has strong psychometric evidence.

A review of the literature reveals that valuable measurement tools on self-efficacy and motivation in the musical field have been developed by McPherson and McCormick (2006) and Ritchie and Williamon (2011). However, these tools generally focus on general musical performance and do not address the specific components required by aptitude tests, such as 'auditory coding (dictation)', 'instant deciphering', and 'jury stress'. The developed scale (the MASES) differs from the

 doi:10.20944/preprints202603.0039.v1

existing literature in that it is the first assessment tool to base the multidimensional structure of aptitude tests on the Triadic Model, which includes Cognitive, Psychomotor, and Affective aspects.

4.1. Validation of the Factor Structure (Triadic Model)

The most important finding of the study is that musical self-efficacy is not a one-dimensional structure, but rather built upon three complementary pillars. The “Cognitive-Auditory” dimension, which emerged in the analyses, represents the student’s belief in their musical perception and theoretical background; while the “Psychomotor” dimension reflects confidence in their technical performance skills. The literature emphasizes that musical performance requires the synchronization of cognitive and motor skills (McPherson & McCormick, 2006; Hallam, 2010). The emergence of the “Affective Regulation” dimension in addition to these two dimensions confirms the critical role of stage anxiety and stress management in exam performance. Considering that performance anxiety, especially during adolescence, directly affects self-efficacy beliefs (Kenny, 2011), the presence of this dimension increases the scale’s comprehensiveness regarding exam psychology. The excellent fit indices obtained in the DFA results ($RMSEA$=.036, CFI=.976) prove that this triadic model works flawlessly in the adolescent sample.

4.2. Reliability and Measurement Accuracy

The reliability analyses of the scale were examined using both classical Cronbach α and modern McDonald’s ω coefficients, and high internal consistency (in the range of .85 - .92) was obtained in all sub-dimensions. The reporting of the ω coefficient is particularly important because it provides more reliable results when the factor loadings of the items are not equal (Hayes & Coutts, 2020). These high reliability values indicate that the MASES can produce stable measurements with low error margins in individual assessments and in creating student profiles before aptitude tests.

4.3. Relationship with General Self-Efficacy (Criterion-Related Validity)

Another important finding of the study is the moderate-to-high level of significant correlation between the MASES and the SEQ-C (r = .586). This finding aligns with Bandura’s (1997) thesis that “Self-efficacy, although domain-specific, is fueled by general competence beliefs.” It is particularly noteworthy that the “Affective Regulation” dimension of the MASES scale shows strong correlations with the “Social” and “Emotional” competency dimensions of the overall scale. This statistically supports the fact that music aptitude tests are not only a process of technical competence but also a process of social and emotional resilience. It can be said that adolescents who feel musically competent also feel more competent in general life skills (or vice versa).

4.4. Limitations and Recommendations

This study is limited to adolescents preparing for the Music Education Department entrance exams of Faculties of Education in Turkey. Future studies are recommended to examine the relationship between students’ MASES scores and actual exam results to test the predictive validity of the scale. Additionally, comparing the “Affective Regulation” dimension with performance anxiety scales (e.g., Kenny Music Performance Anxiety Inventory) could contribute to the discriminant validity of the scale.

5. Conclusions and Recommendations

5.1. Conclusions

This research encompasses the development process of the Music Aptitude Self-Efficacy Scale (MASES), which aims to measure the musical aptitude self-efficacy perceptions of adolescents (16-18 years old) preparing for music aptitude tests in a multidimensional way. Based on the psychometric findings, the following main conclusions were reached:

Structural Validity; The scale consists of three core factors: (1) Psycho-Motor Performance Competence, (2) Cognitive-Auditory Competence, and (3) Affective Regulation, in line with theoretical models in the literature (Triadic Model). Confirmatory Factor Analysis (CFA) results (*RMSEA* = .036, *CFI* = .976) demonstrated that this three-factor structure works perfectly in the adolescent sample.

Reliability; The high internal consistency (Cronbach α) and composite reliability (McDonald's ω and CR) coefficients (in the range of .85 - .93) for all sub-dimensions indicate that the MASES produces stable and reliable results free from measurement error.

Discrimination and Association; the MASES demonstrated a positive association with general competence beliefs (the SEQ-C) at the expected level (r = .586), thus establishing its criterion-related validity. It has also revealed that musical self-efficacy is a domain-specific construct, distinct from general academic or social competence.

In conclusion, the MASES is a standardized assessment tool for music educators and examination juries, allowing them to measure not only candidates' musical capacity but also their belief in translating that capacity into performance (Appendix B). This scale will enable the identification of students with high anxiety levels or low self-efficacy before the talent exam, and provide them with guidance and support.

5.2. Recommendations

Based on the research findings, the following recommendations have been developed for practitioners, educators, and researchers:

5.2.1. Recommendations for Practitioners, and Educators

Diagnostic Assessment: Educators working in the guidance services of Fine Arts High Schools and Conservatories, as well as those in exam preparation courses, can use MASES to identify students' "risky" areas at the beginning of the educational process. For example, a student with a high "Psychomotor" score but a low "Affective Regulation" score, even if technically proficient, is at risk of failure due to exam anxiety. Performance simulations and anxiety management exercises should be prioritized for these students.

Individualized Guidance: School guidance services can utilize this scale to determine the career orientation of students preparing for aptitude tests. Students with consistently low self-efficacy perceptions can be supported in restructuring their expectations and goals.

Profiling for Exam Juries: Aptitude test juries can include the MASES scores as "complementary data" in the evaluation process (e.g., interview stage) to assess their belief in candidates' potential beyond their immediate performance.

5.2.2. Recommendations for Researchers:

Predictive Validity: Future studies should examine the relationship between students' pre-exam the MASES scores and their actual placement scores. Testing the scale's predictive power on exam success (pass/fail) will make a significant contribution to the literature.

Longitudinal Studies: To observe changes in musical self-efficacy over time, repeated measurements can be taken at the beginning, middle, and on the day of the exam.

Cross-Cultural Adaptation: To ensure the scale is not limited to the Turkish exam system, its cross-cultural validity can be tested by adapting it to students preparing for similar music aptitude tests (ABRSM, Conservatory entrance exams, etc.) in different countries.

Relationship with Performance Anxiety: Examining the relationship between the "Affective Regulation" dimension of the MASES and specific anxiety scales such as the Kenny Music Performance Anxiety Inventory (K-MPAI) will help understand the theoretical balance between self-efficacy and anxiety.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1, S2 & S3; Appendix A, Appendix B.

Author Contributions: Conceptualization, G.C.Ş., A.K. and Ç.B.; methodology, G.C.Ş., A.K. and Ç.B.; software, A.K.; validation, G.C.Ş. and A.K.; formal analysis, G.C.Ş.; investigation, Ç.B.; resources, G.C.Ş., A.K., Ç.B.; data curation, A.K.; writing—original draft preparation, G.C.Ş., A.K., Ç.B.; writing—review and editing, G.C.Ş.; visualization, A.K. and Ç.B.; supervision, G.C.Ş.; project administration, A.K.; funding acquisition, G.C.Ş., A.K. and Ç.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Mugla Sitki Kocman University (protocol code: 162, and Date of Approval: April 28, 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data related to the study have been submitted to the journal in the article.

Acknowledgments: During the preparation of this article/study, the author(s) used Gemini 3.1 Pro for visualization (Figure 1) and proofreading. The authors have reviewed and edited the output and are fully responsible for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MASES	The Music Aptitude Self- Efficacy Scale
CFA	Confirmatory Factor Analysis
EFA	Exploratory Factor Analysis

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411. https://doi.org/10.1037/0033-2909.103.3.411

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Prentice-Hall.

Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological methods & research*, *16*(1), 78-117. https://doi.org/10.1177/0049124187016001004

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* David McKay. https://doi.org/10.1177/001316445701700420

Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford publications.

Browne, M. W., Cudeck, R., Bollen, K. A., & Long, J. S. (1993). Alternative ways of assessing model fit. *Testing structural equation models*, *154*(4), 136-162.

Büyüköztürk, Ş. (2011). Sosyal bilimler için veri analizi el kitabı. Pegem Akademi.

Byrne, B. M. (2010). Structural equation modeling with AMOS: Basic concepts, applications, and programming (2nd ed.). Routledge.

Celikkaleli, Ö., Gündoğdu, M., & Kıran-Esen, B. (2006). Ergenlerde yetkinlik beklentisi ölçeğinin Türkçe uyarlaması: Geçerlik ve güvenirlik çalışması. *Eğitim Araştırmaları Dergisi*, *25*(1), 62-72. https://search.trdizin.gov.tr/en/yayin/detay/64566

Christiner, M., Bernhofs, V., Sommer-Lolei, S., & Groß, C. (2023). What Makes a Foreign Language Intelligible? An Examination of the Impact of Musical Ability and Individual Differences on Language Perception and How Intelligible Foreign Languages Appear. *Journal of Intelligence*, *11*(3), 43. https://doi.org/10.3390/jintelligence11030043

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis.* Psychology press.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation, 10*(1). https://doi.org/10.7275/jyj1-4868

Creswell J. W. (2013). Steps in conducting a scholarly mixed methods study. DBER Speaker Series.

Creswell, J. W., & Clark, V. P. (2007). *Mixed methods research.* Thousand Oaks, CA.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458-467. https://doi.org/10.1287/mnsc.9.3.458

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, *5*(4), 194-197. https://doi.org/10.1016/S0897-1897(05)80008-4

DeVellis, R. F. (2017). *Scale development: Theory and applications.* Sage publications.

Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of educational measurement.* Englewood Cliffs: Prentice-Hall.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

George, D., & Mallery, P. (2019). IBM SPSS statistics 26 step by step: A simple guide and reference (16th ed.). Routledge.

Gordon, E. E. (2012). Learning sequences in music: A contemporary music learning theory. GIA Publications.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hallam, S. (2010). The power of music: Its impact on the intellectual, social and personal development of children and young people. *International Journal of Music Education*, *28*(3), 269-289. https://doi.org/10.1177/0255761410370658

Hambrick, D. Z., Burgoyne, A. P., Altmann, E. M., & Matteson, T. J. (2023). Explaining the Validity of the ASVAB for Job-Relevant Multitasking Performance: The Role of Placekeeping Ability. *Journal of Intelligence*, *11*(12), 225. https://doi.org/10.3390/jintelligence11120225

Hauenstein, C. E., Embretson, S. E., & Kim, E. (2024). Psychometric Modeling to Identify Examinees' Strategy Differences during Testing. *Journal of Intelligence*, *12*(4), 40. https://doi.org/10.3390/jintelligence12040040

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *Communication Methods and Measures*, *14*(1), 1-24. https://doi.org/10.1080/19312458.2020.1718629

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, *7*(2), 191-205. https://doi.org/10.1177/1094428104263675

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kenny, D. T. (2011). *The psychology of music performance anxiety*. Oxford University Press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Publications.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: Handbook II: Affective domain.* David McKay.

Krueger, R. A., & Casey, M. A. (2015). Focus group interviewing. *Handbook of practical program evaluation*, 506-534.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications.* Addison-Wesley.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research,* 35(6), 382-385. https://journals.lww.com/nursingresearchonline/citation/1986/11000/determination_and_quantification_of_content.17.aspx

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling*, *11*(3), 320-341. https://doi.org/10.1207/s15328007sem1103_2

McPherson, G. E., & McCormick, J. (2006). Self-efficacy and music performance. *Psychology of Music*, *34*(3), 322-336. https://doi.org/10.1177/0305735606064841

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis*. Sage.

Morgan, D. L. (1997). *Focus groups as qualitative research* (Vol. 16). Sage.

Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. *Journal of Psychopathology and Behavioral Assessment*, *23*(3), 145-149. https://doi.org/10.1023/A:1010961119608

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory 3rd edition*. MacGraw-Hill, New York.

Patton, M. Q. (2014). Qualitative research & evaluation methods: Integrating theory and practice. Sage publications.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. https://doi.org/10.1002/nur.20147

Ritchie, L., & Williamon, A. (2012). Self-efficacy as a predictor of musical performance quality. *Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 334-340. https://doi.org/10.1037/a0029619

Sideridis, G., & Alahmadi, M. (2022). Estimation of Person Ability under Rapid and Effortful Responding. *Journal of Intelligence*, *10*(3), 67. https://doi.org/10.3390/jintelligence10030067

Simpson, E. J. (1972). The classification of educational objectives in the psychomotor domain. Gryphon House.

Strauss, H., Reiche, S., Dick, M. et al. (2024). Online assessment of musical ability in 10 minutes: Development and validation of the Micro-PROMS. *Behav Res* 56, 1968–1983. https://doi.org/10.3758/s13428-023-02130-4

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (Vol. 4). Mahwah, NJ: Lawrence erlbaum associates.

Sust, L., Bergmann, M., Bühner, M., & Schoedel, R. (2026). Deep Beats, Deep Thoughts? Predicting General Cognitive Ability from Natural Music-Listening Behavior. *Journal of Intelligence*, *14*(2), 29. https://doi.org/10.3390/jintelligence14020029

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tunca, O. E., Kumtepe, E. G., Torun, S., & Can Ugurhan, Y. Z. (2024). Development and Validation of The Anadolu University Music Perception Test. *International Journal of Music Education*. https://doi.org/10.1177/02557614231193422

Willis, G. B. (2005). Cognitive interviewing: A tool for improving questionnaire design. Sage Publications.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806-838. https://doi.org/10.1177/0011000006288127

Yıldırım, A., & Şimşek, H. (2018). *Sosyal bilimlerde nitel araştırma yöntemleri.* Seçkin Yayıncılık.

Zelenak, M. S. (2015). Measuring the Sources of Self-Efficacy Among Secondary School Music Students. *Journal of Research in Music Education*, 62(4), 389-404. https://doi.org/10.1177/0022429414555018

Zelenak, M. S. (2024). Self-efficacy and music performance: A meta-analysis. *Psychology of Music*, *52*(6), 649-667. https://doi.org/10.1177/03057356231222432

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91. https://doi.org/10.1006/ceps.1999.1016