An Empirical Study on the Effectiveness of Digital Tools Applied in Adolescent Music Aural Training

Junge Chang*

Russian State Pedagogical University, Saint Petersburg, Russia

*Email: changjunge2027@163.com

Abstract. Empirical evidence on whether digital tools effectively support adolescent music aural training remains scarce. This study employed a 12-week quasi-experimental pre-test/post-test design with 120 adolescents (aged 12–17) randomly assigned to an experimental group trained with adaptive digital software (EarMaster Cloud) or a control group receiving equivalent teacher-led instruction. Performance in pitch, rhythm, interval, and chord perception was measured, and data were analyzed via repeated-measures ANOVA, MANOVA, and hierarchical regression. Results showed that digital-tool training significantly improved overall aural skills, with effects varying across dimensions and moderated by prior musical experience and engagement. Findings support digital tools as complementary—rather than substitutive—resources, most effective for pitch and interval perception while teacher-led training remains essential for harmonic skills.

Keywords: Digital tools, Adolescent learners, Music aural training, Solfeggio and ear-training, Empirical effectiveness study

1 Introduction

Digital technologies have reshaped instructional practice across disciplines requiring sustained practice, including music. Aural competence—pitch and rhythm discrimination, interval and harmony identification—has traditionally been cultivated through teacher-led drill, but adaptive software, intelligent tutoring systems, and mobile applications are increasingly moving from supplementary to central roles in music learning. Bibliometric analyses confirm that educational technology in music pedagogy has become one of the fastest-growing research areas since 2020[1].

Despite this growth, empirical evidence on digital aural training for adolescents remains thin. Most existing research targets instrumental instruction or university-level learners [2-3], leaving adolescents—whose auditory perception and musical literacy are highly plastic—largely unexamined. The present study addresses this gap by testing (a) whether digital tools improve adolescents' overall aural competence, (b) whether effects differ across skill dimensions (pitch, rhythm, interval, chord), and (c) how prior musical experience and engagement moderate outcomes.

T. Chamaratana et al. (eds.), *Proceedings of the 2026 12th International Conference on Digital Humanities and Frontiers in Social Sciences (DHFSS 2026)*, Advances in Social Science, Education and Humanities Research 1031,
https://doi.org/10.2991/978-2-38476-611-6_26

Three hypotheses are tested. H1: Digital-tool training significantly improves overall aural competence. H2: Effect magnitudes vary across dimensions, being largest for pitch and interval and smallest for chord discrimination. H3: Greater prior musical experience and engagement yield larger effects.

2 Literature Review

2.1 Definition of Core Concepts

Digital technologies in music instruction here refer to software applications, mobile apps, intelligent tutoring systems, and web-based platforms designed to support aural skill acquisition. Aural training is the cultivation of listening-based skills—pitch, rhythm, interval, chord, melodic dictation, harmony—which modern research treats as integrated temporal-spectral processing rather than isolated subskills [4]. Adolescent learners, defined here as ages 12–18, represent a developmentally distinct group whose auditory, cognitive, and musical capacities are simultaneously maturing [5]; findings from children or adults cannot be assumed to transfer.

2.2 Review of International Studies

Recent international research has concentrated on creative composition rather than auditory perception, with sequencers and platforms such as GarageBand dominating K–12 applications. A second strand examines gamified digital learning, finding that while engagement improves, clear evidence of skill gains has yet to be established. A third strand evaluates AI-assisted sight-singing and aural training in higher education, reporting significant gains in pitch, rhythm, and melodic dictation under constructivist or hybrid frameworks [6]—but the tertiary sample limits generalizability to adolescents.

2.3 Review of Domestic Studies

Chinese empirical research has grown since 2020. Lingjie [7] reported significant improvements in pitch, rhythm, and timbre perception among college students using EarMaster, validating pre-/post-test designs but again leaving adolescents unaddressed. Jiang and Zhang [8] found that perceived effectiveness, social influence, and facilitating conditions predict teachers' technology integration more strongly than ease of use, underscoring the need for empirical effectiveness data.

2.4 Research Commentary and Entry Point of This Study

Three gaps justify the present study: (1) aural training is underrepresented relative to creativity and instrumental research; (2) adolescents are rarely studied as a distinct group; (3) individual difference moderators (prior experience, engagement) are seldom modeled. This study addresses all three through a multi-dimensional, moderated empirical design focused on adolescents.

3 Data and Methods

3.1 Research Participants

Participants were recruited from three secondary schools and two music institutes in eastern urban China. From 128 screened, 120 adolescents (60 male, 60 females; M age = 14.6, SD = 1.42) meeting the inclusion criteria (≥1 year of formal music study, no auditory disorders) were retained. They were stratified by prior musical training (1–2, 3–4, or ≥5 years) and randomly assigned to experimental (digital-tool) or control (teacher-led) groups (n = 60 each). As shown in Table 1, the two groups were well matched in age, gender distribution, prior musical training background, and school type, ensuring satisfactory baseline equivalence for subsequent experimental analysis.

Table 1. Demographic Characteristics of Study Participants by Group Assignment.

Characteristic	Experimental Group (n = 60)	Control Group (n = 60)	Total Sample (N = 120)
Age (years)			
Mean (SD)	14.5 (1.39)	14.7 (1.45)	14.6 (1.42)
Range	12 – 17	12 – 17	12 – 17
Gender, n (%)			
Male	30 (50.0)	30 (50.0)	60 (50.0)
Female	30 (50.0)	30 (50.0)	60 (50.0)
Prior Musical Training, n (%)			
Beginners (1–2 years)	21 (35.0)	22 (36.7)	43 (35.8)
Intermediate (3–4 years)	23 (38.3)	22 (36.7)	45 (37.5)
Advanced (≥ 5 years)	16 (26.7)	16 (26.7)	32 (26.7)
School Type, n (%)			
Secondary school	38 (63.3)	36 (60.0)	74 (61.7)
Music training institution	22 (36.7)	24 (40.0)	46 (38.3)

Note: The randomization procedure produced groups with comparable distributions across all demographic variables, supporting baseline equivalence for subsequent inferential analyses.

3.2 Research Instruments

EarMaster Cloud served as the experimental digital tool, chosen for its modular coverage of pitch, intervals, chords, rhythm, and melodic dictation and its adaptive immediate-feedback design [9]. Aural competence was measured via a 60-item battery (15 items per construct: pitch, rhythm, interval, chord), piloted with 20 non-sample adolescents; Cronbach's α = 0.87. Learner experience was assessed via an 18-item, 5-point Likert questionnaire (perceived usefulness, ease of use, engagement, satisfaction); Cronbach's α = 0.91.

3.3 Research Design

A quasi-experimental pre-test/mid-test/post-test design was used over 12 weeks (24 hours of instruction, two 1-hour sessions per week). The experimental group used

EarMaster Cloud under teacher supervision; the control group received equivalent teacher-led instruction with identical curriculum coverage. Pre-test occurred in Week 1, mid-test in Week 7, and post-test in Week 12. This multi-method design addresses a methodological concern raised in systematic evaluations of digital technologies in music education that effects are often inferred from a single data source [10]. The overall research procedure and timeline are illustrated in Fig. 1.

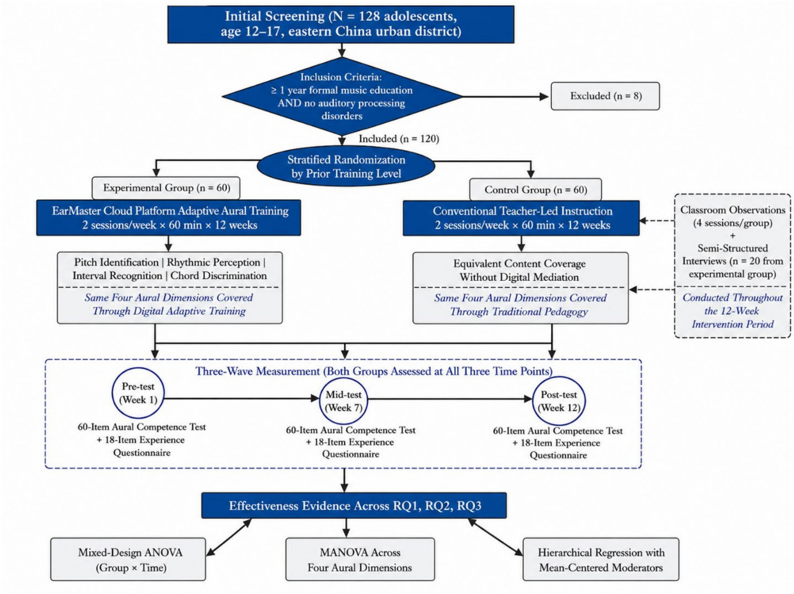

Fig. 1. Research Design and Procedural Flow of the 12-Week Intervention Study.

3.4 Data Collection Procedures

Data were collected at three time points to enable both cross-sectional and longitudinal analysis. The pre-test was administered in Week 1 under standard classroom conditions, comprising the 60-item aural competence test and the 18-item learner experience questionnaire. The mid-test was administered in Week 7 to capture mid-intervention progress, and the post-test in Week 12. Structured classroom observations were conducted in four randomly selected classes per group, with two trained observers recording engagement levels, fidelity of curriculum implementation, and learner behavior; observations served exclusively as triangulating evidence for the quantitative data. In addition, 20 participants from the experimental group—stratified across the three prior-training levels—were interviewed for 25–30 minutes in the final week.

3.5 Data Processing and Analytical Methods

Quantitative analyses were performed in SPSS 27.0 with α = 0.05. Descriptive statistics (means, standard deviations, 95% confidence intervals) were computed for all variables to characterize the sample and identify outliers prior to inferential analysis. Baseline equivalence between groups on pre-test scores was assessed using independent-samples t-tests.

Hypothesis 1 was tested using a two-way mixed ANOVA, with group as the between-subjects factor and time as the within-subjects factor. The practical significance of the time effect was quantified using partial eta squared, computed as shown in Equation (1):

$$\eta_p^2 = \frac{SS_{time}}{SS_{time} + SS_{error}} \tag{1}$$

Partial eta squared was selected because, in mixed designs, it partitions variance into theoretically meaningful components and supports the comparability of effect sizes across the model's terms.

Hypothesis 2 was tested using MANOVA across the four aural subskills (pitch, rhythm, interval, chord), with univariate follow-up tests. MANOVA was chosen over multiple separate ANOVAs to control Type I error inflation while accounting for correlations among the subskills.

Hypothesis 3—concerning the moderating role of individual differences—was tested using hierarchical regression on aural competence gain scores (post-test minus pre-test). All continuous moderators were mean-centered prior to estimation to reduce multicollinearity between main effects and interaction terms. The full regression model is specified in Equation (2):

$$Y_{gain} = \beta_0 + \beta_1 X_{group} + \beta_2 X'_{prior} + \beta_3 X'_{engagement} + \beta_4 (X_{group} \times X'_{prior}) + \beta_5 (X_{group} \times X'_{engagement}) + \varepsilon \tag{2}$$

where Training_c and Engagement_c denote the mean-centered moderators. The hierarchical structure (Block 1: group; Block 2: moderators; Block 3: interactions) provides a conservative test of incremental explanatory power and isolates the moderation effects predicted under H3.

Qualitative interview data were analyzed thematically along three a priori themes—training effectiveness, engagement, and challenges encountered. Inter-rater agreement between two independent coders was assessed using Cohen's $\kappa = 0.84$. Integrating statistical and qualitative analyses provides the mixed-methods triangulation strategy required to address methodological concerns in this research area.

4 Results

4.1 Overall Changes in Aural Competence Before and After Training

Baseline equivalence was confirmed, $t(118) = 0.31$, $p = .755$. As shown in Table 2, the experimental group's composite scores rose from $M = 38.42$ ($SD = 6.18$) to $M = 53.18$ ($SD = 5.84$), a gain of 14.76 points, while the control group's rose from $M = 38.06$ ($SD = 6.31$) to $M = 44.36$ ($SD = 5.93$), a gain of 6.30 points. The two-way mixed ANOVA produced a significant Group × Time interaction, $F(2, 236) = 58.47$, $p < .001$, $\eta^2_p = 0.331$, supporting H1. The changing trajectory of aural competence across three test time points for both groups is visualized in Fig. 2.

Table 2. Descriptive Statistics of Aural Competence Composite Scores Across Three Measurement Occasions by Group.

Measurement Occasion	Experimental Group (n = 60)			Control Group (n = 60)		
	M	SD	95% CI	M	SD	95% CI
Pre-test	38.42	6.18	[36.83, 40.01]	38.06	6.31	[36.43, 39.69]
Mid-test (Week 7)	46.91	5.97	[45.37, 48.45]	41.27	6.12	[39.69, 42.85]
Post-test (Week 12)	53.18	5.84	[51.67, 54.69]	44.36	5.93	[42.83, 45.89]
Gain (Post – Pre)	14.76	—	—	6.30	—	—

Note: Scores on a 0–60 scale. Baseline $t(118) = 0.31$, $p = .755$. Group × Time interaction: $F(2, 236) = 58.47$, $p < .001$, $\eta^2_p = 0.331$.

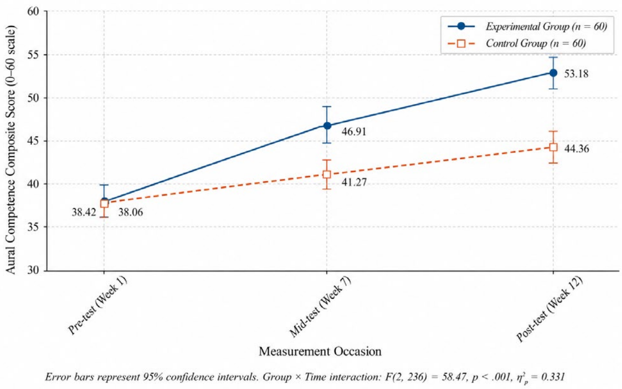

Fig. 2. Trajectory of Aural Competence Composite Scores Across Pre-test, Mid-test, and Post-test by Group.

4.2 Effectiveness Analysis Across Different Aural Dimensions

MANOVA on the four subskills yielded a significant overall group effect, Wilks' $\Lambda = 0.561$, $F(4, 115) = 22.48$, $p < .001$, $\eta^2_p = 0.439$. Univariate tests confirmed significant Group × Dimension differentials across all four (all $p < .01$). As shown in Table 3, effect sizes for the experimental group followed a clear hierarchy: pitch identification ($d = 1.62$) > interval recognition ($d = 1.43$) > rhythmic perception ($d = 1.24$) > chord discrimination ($d = 0.97$). Control-group gains ranged $d = 0.31$–0.54. The pattern supports H2: digital tools differentially benefit pitch and interval perception more than chord discrimination, consistent with the cognitive complexity of harmonic processing.

Table 3. Dimensional Analysis of Training Gains Across Four Aural Skill Domains.

Aural Dimension	Experimental Group (n = 60)				Control Group (n = 60)			
	Pre-test M (SD)	Post-test M (SD)	Gain	Cohen's d	Pre-test M (SD)	Post-test M (SD)	Gain	Cohen's d
Pitch Identification	9.83 (1.74)	13.74 (1.58)	3.91	1.62	9.71 (1.79)	11.18 (1.71)	1.47	0.54
Rhythmic Perception	9.92 (1.81)	13.21 (1.65)	3.29	1.24	9.86 (1.85)	11.05 (1.74)	1.19	0.42
Interval Recognition	9.45 (1.72)	13.07 (1.63)	3.62	1.43	9.38 (1.76)	10.84 (1.69)	1.46	0.51
Chord Discrimination	9.22 (1.83)	12.16 (1.78)	2.94	0.97	9.11 (1.86)	10.27 (1.81)	1.16	0.31

Note: Each subskill scored 0–15. MANOVA: Wilks' $\Lambda = 0.561$, $F(4, 115) = 22.48$, $p < .001$, $\eta^2_p = 0.439$.

4.3 Influence of Individual Difference Variables

Hierarchical regression (Table 4) showed group assignment alone explained 28.7% of gain-score variance. Adding moderators raised R^2 to 0.423; adding interactions raised it to 0.486 ($\Delta R^2 = 0.063$, $F(2, 114) = 7.01$, $p = .001$). Both Group × Prior Training ($\beta = 0.218$, $p = .005$) and Group × Engagement ($\beta = 0.247$, $p = .002$) interactions were significant, supporting H3. Simple-slope analyses within the experimental group revealed gains of 10.84, 14.62, and 17.21 points for beginners, intermediates, and advanced learners, respectively—indicating that digital-tool benefits scale with prior preparation and engagement. The simple slope patterns across different prior training levels are presented in Fig. 3.

Table 4. Hierarchical Regression Results Predicting Aural Competence Gain Scores.

Predictor	Block 1		Block 2		Block 3	
	β	t	β	t	β	t
Group Assignment	0.536***	6.91	0.512***	7.04	0.498***	6.87
Prior Training (centered)	—	—	0.281***	3.86	0.263***	3.62
Engagement (centered)	—	—	0.317***	4.35	0.294***	4.07
Group × Prior Training	—	—	—	—	0.218**	2.86
Group × Engagement	—	—	—	—	0.247**	3.21
Model Statistics						
R^2	0.287		0.423		0.486	
ΔR^2	0.287		0.136		0.063	
F for ΔR^2	47.43***		13.78***		7.01**	
df	(1, 118)		(2, 116)		(2, 114)	

Note: DV = gain score (post – pre). Continuous predictors mean-centered. Group coded 0 = control, 1 = experimental. ** $p < .01$, *** $p < .001$.

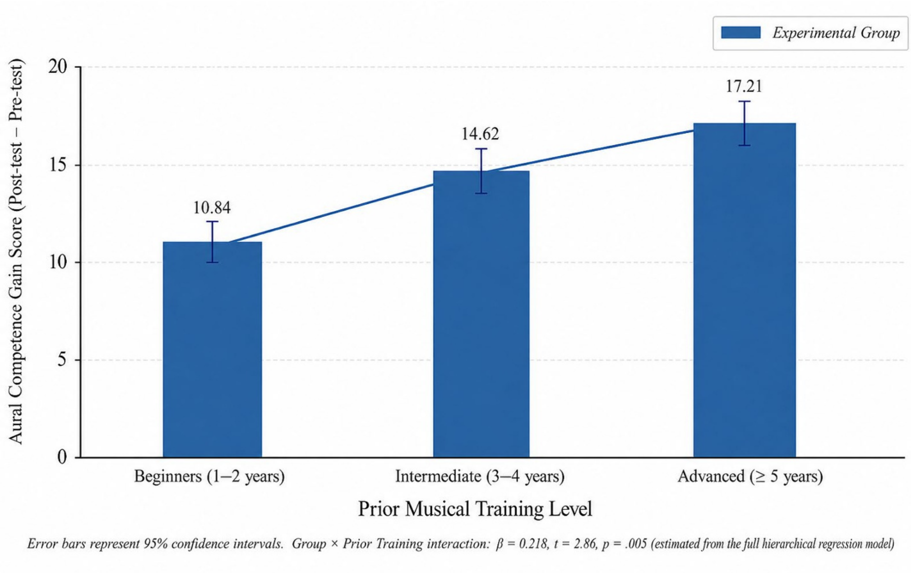

Fig. 3. Simple Slope Analysis of Gain Scores by Prior Training Level Within the Experimental Group.

4.4 Learners' User Experience and Attitudinal Feedback

Learner-experience ratings (Table 5) were positive across all four constructs (range 3.91–4.34). Engagement scored highest (M = 4.34) and ease of use lowest (M = 3.91), with the larger SD on ease of use suggesting variability tied to digital literacy. Interview thematic analysis surfaced three themes: (1) immediate feedback—valued by 16 of 20 participants, though three noted limited contextual explanation; (2) initial cognitive overload from interface multitasking, resolving after 3–4 weeks for most; (3) gamified motivation—sustained for 11 participants throughout the 12 weeks, but waning for four.

Table 5. Learner Experience Questionnaire Results in the Experimental Group.

Dimension	Number of Items	M	SD	95% CI	Cronbach's α
Perceived Usefulness	5	4.27	0.61	[4.11, 4.43]	0.88
Ease of Use	4	3.91	0.74	[3.72, 4.10]	0.83
Engagement	5	4.34	0.58	[4.19, 4.49]	0.89
Satisfaction	4	4.18	0.66	[4.01, 4.35]	0.85
Overall	18	4.18	0.65	[4.01, 4.35]	0.91

Note: 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Administered following the Week 12 post-test.

5 Discussion

The experimental group's 38% gain in aural competence cannot be attributed to novelty alone, as the upward trajectory continued from mid-test to post-test rather than decaying. Three pedagogical mechanisms appear central: immediate feedback, adaptive task difficulty, and spaced repetition—consistent with meta-analytic findings that adaptive feedback drives gains in pitch, rhythm, and auditory retention [11]. The present study extends these mechanisms to adolescent learners.

The dimensional hierarchy (pitch > interval > rhythm > chord) aligns with prior evidence on temporal and pitch perception in children, but the gap between pitch and chord effects here is wider than predicted. This suggests digital tools may produce disproportionate skill profiles favoring melodic over harmonic competence, with implications for curriculum balance.

Moderation findings indicate the technology benefits experienced learners disproportionately—novice/intermediate/advanced gaps of 5.41, 6.78, and 7.43 points (control) widen to 10.84, 14.62, and 17.21 (experimental). Combined with Jiang and Zhang's [8] finding that institutional adoption of educational technology depends primarily on demonstrated effectiveness rather than ease of use, this underscores the need for empirical evidence such as the present study and supports differentiated deployment: intensive digital training for advanced learners, integrated approaches for beginners.

Limitations on chord discrimination, early cognitive load, and waning gamified motivation indicate digital tools are a valuable but bounded resource. Consistent with critical reviews of AI-assisted music education, structural limits—accessibility, transparency, algorithmic opacity—warrant caution. Practically, digital tools are best deployed for pitch and interval training, with teacher-led instruction retained for harmony.

6 Conclusion

This 12-week study of 120 adolescents provides empirical evidence that digital-tool aural training significantly improves overall aural competence relative to teacher-led instruction (38% vs. lower gains), with the largest effects in pitch identification and the smallest in chord discrimination. Prior musical experience and engagement function as significant but modest moderators (accounting for 6.3% additional variance), with the digital tool remaining the primary driver of gains.

Three contributions follow: focusing empirically on adolescents as a distinct population; adopting a multidimensional assessment across pitch, rhythm, interval, and chord; and modeling moderation by individual learner characteristics. Limitations include the 12-week duration (precluding inferences about long-term retention), the single-tool design (EarMaster Cloud only), and the geographically constrained sample (eastern urban China). Future work should pursue longitudinal designs across academic years, comparative evaluation of multiple platforms, cross-regional replication, and neurocognitive measures of underlying perceptual and attentional processes.

References

1. Ma, Y., Wang, C. Empowering music education with technology: a bibliometric perspective. Humanit Soc Sci Commun 12, 345 (2025). https://doi.org/10.1057/s41599-025-04616-2.
2. Acquilino A and Scavone G (2022) Current State and Future Directions of Technologies for Music Instrument Pedagogy. Front. Psychol. 13:835609. doi: 10.3389/fpsyg.2022.835609.
3. He, L., and Z.-l. Chen. 2026. “AI-Powered Auditory Training and Its Role in Strengthening Musical Memory: A Systematic Review in Higher Music Education.” European Journal of Education61, no. 2: e70569. https://doi.org/10.1111/ejed.70569.
4. Psarris, G., Eleftheriadis, N., Sidiras, C., Sereti, A., & Iliadou, V. M. (2024). Temporal resolution and pitch discrimination in music education: Novel data in children. European Archives of Oto-Rhino-Laryngology, 281(8), 4103-4111. https://doi.org/10.1007/s00405-024-08571-7.
5. Müllensiefen, D., Elvers, P., & Frieler, K. (2022). Musical development during adolescence: Perceptual skills, cognitive resources, and musical training. Annals of the New York Academy of Sciences, 1518(1), 264-281. https://doi.org/10.1111/nyas.14911.
6. Chen, Z. L., & Lin, H. (2026). AI and digital technologies in sight-singing and aural training: A systematic review of innovations in higher music education. European Journal of Education, 61(1), e70502. https://doi.org/10.1111/ejed.70502.
7. Lingjie, A. (2025). Effectiveness study of applying EarMaster for ear training in Chinese university students. International Journal of Sociologies and Anthropologies Science Reviews, 5(2), 29-38. DOI:10.60027/ijsasr.2025.5514.
8. Jiang, J., & Zhang, X. (2025). Chinese music teachers value effectiveness, social, and facilitating factors over ease of use in ICT integration: A PLS-SEM study. PLOS ONE, 20(11), e0334186. https://doi.org/10.1371/journal.pone.0334186.
9. Ou, J., Nogueira, J., & Qin, C. (2025). Exploring the impact of AI-assisted practice applications on music learners' performance, self-efficacy, and self-regulated learning. Frontiers in Psychology, 16, 1675762. doi: 10.3389/fpsyg.2025.1675762.
10. Rexhepi, F. G., Breznica, R. K., & Rexhepi, B. R. (2024). Evaluating the effectiveness of using digital technologies in music education. Journal of Educational Technology Development and Exchange, 17(1), 273-289. https://doi.org/10.18785/jetde.1701.16.
11. He, L., & Chen, Z. L. (2026). AI-powered auditory training and its role in strengthening musical memory: A systematic review in higher music education. European Journal of Education, 61(2), e70569. https://doi.org/10.1111/ejed.70569.