

Understanding music teachers' formative assessment intention and implementation: A Chinese Mainland context

International Journal of
Music Education
1–19

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/02557614251315461

journals.sagepub.com/home/ijm



Le-Xuan Zhang  and **Zi Yan** 
The Education University of Hong Kong, Hong Kong

Xiang Wang

Hong Kong Baptist University, Hong Kong

Abstract

Formative assessment is essential in music education as it supports the music learning process, which relies mostly on feedback from self and others to improve performance. Despite growing interest in formative assessment across various subjects, there is a lack of empirical evidence on how it is applied specifically in music education. Given the crucial role of teachers in the effective implementation of formative assessment, this study aims to examine music teachers' intentions and implementation of formative assessment, along with the factors influencing them, based on the Theory of Planned Behavior. A total of 671 music teachers from 29 cities/provinces of Mainland China were surveyed. The structural equation modeling results indicate that in the Chinese Mainland school music education context, a positive attitude toward formative assessment, a supportive and collaborative social environment, and strong confidence in using formative assessment enhance teachers' intentions to adopt it. Additionally, greater confidence directly increases actual implementation. However, increased school support did not significantly impact teachers' intentions or their implementation of formative assessment. The findings suggest that school administrators should focus on helping music teachers build confidence and fostering a collaborative, supportive culture for using formative assessment practices to improve music learning.

Keywords

China, formative assessment, music education, music teacher, theory of planned behavior

Corresponding author:

Zi Yan, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong.

Email: zyan@eduhk.hk

Introduction

Formative assessment has gained global attention across various subjects due to its advantages in monitoring student progress, offering feedback, and addressing individual learning needs (Black & Wiliam, 2009; Yan et al., 2021). In the context of Chinese Mainland music education, formative assessment is promoted in policies and curriculum documents to support its implementation in schools for the benefit of students' learning (Ministry of Education, 2001). However, after two decades of promotion, the implementation of formative assessment in school settings remains unsatisfactory. This is primarily attributed to teachers' reluctance to adopt new assessment practices and a limited understanding of the factors that could facilitate or impede their formative assessment practices (Yan & Cheng, 2015; Yan et al., 2021). Notably, there has been a lack of efforts to gather large-scale, quantitative data to offer a holistic overview of factors affecting music teachers' intentions and implementation regarding formative assessment. This highlights a gap in effectively translating formative assessment from theory to practice and underscores the urgent need for comprehensive quantitative studies.

This study aims to investigate formative assessment in the context of mainland Chinese music teaching, particularly focusing on teachers' intentions and implementation of formative assessment in schools. As both personal and contextual factors are likely to influence these practices (Yan et al., 2021), the study adopts the extended Theory of Planned Behavior (TPB), covering both personal factors and contextual factors, to quantitatively examine the relationships among teachers' intentions, practices, and influencing factors.

Formative assessment in Chinese Mainland music education

Formative assessment, as defined in this study, is a student-centered approach (SCE) that emphasizes ongoing evaluation and feedback to improve learning and teaching. According to Black and Wiliam (1998), it involves a wide range of activities that provide information to be used as feedback for modifying future teaching and learning strategies. These activities can be carried out by teachers (e.g., questioning and feedback) and/or by students themselves (e.g., peer and self-assessment).

Unlike summative assessment, which summarizes learning outcomes, formative assessment focuses on feedback to deepen understanding and develop skills (Brown, 2019). It helps teachers track progress, address learning issues, and support students in reflecting on their learning and enhancing self-regulation (Booth & Kinsella, 2022). Given these benefits, many national curricula have adopted formative assessment across various subjects.

In the context of Chinese Mainland music education, the concept of formative assessment (Chinese: *xing cheng xing ping jia*) emerged in the mid-1980s (Lu, 1987; S. Yuan & Shu, 2017). Initially, it was introduced to counterbalance the summative, examination-driven system by tracking students' progressive learning in music. In 2001, formative assessment was officially incorporated into the school music curriculum to align with the broader goals of Student-Centered Education (SCE), specifically to enhance students' awareness of their own learning and foster greater autonomy (Ministry of Education, 2001). However, despite two decades of promotion, the implementation of formative assessment in Chinese schools faces significant challenges. Research identifies two key limitations that hinder its effective adoption.

First, while formative assessment includes various activities, such as peer assessment and self-assessment, these practices require additional time and effort to implement (Yan & Cheng, 2015). Given music teachers' heavy workloads and limited teaching hours, they are often reluctant to invest extra time in learning and applying new teaching approaches, even if they are shown to be

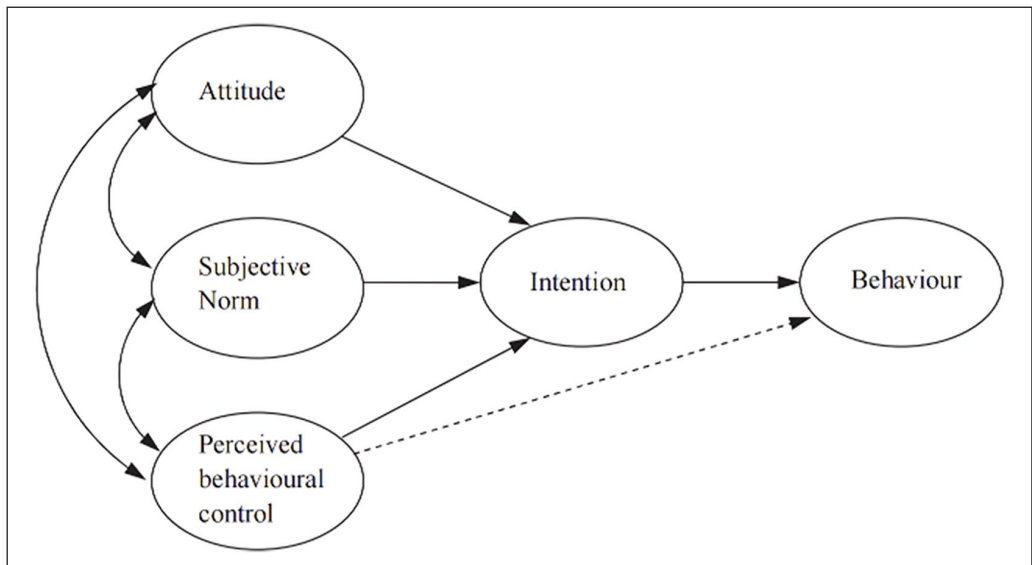


Figure 1. Model of the theory of planned behavior (Ajzen, 1991).

beneficial for both teaching and learning (Zhang & Leung, 2023). As a result, formative assessment in practice relies heavily on teacher-directed feedback, with peer and self-assessment rarely observed in both Chinese demonstration lessons and regular classes (Zhang et al., 2023; Zhang & Leung, 2024). This hesitance not only limits the broader implementation of formative assessment but also undermines the core SCE objective of fostering student autonomy and self-awareness.

Second, much of the existing research on formative assessment, both in Western contexts (Booth & Kinsella, 2022; Kordeš et al., 2014) and in China (Deng, 2021; Z. Yuan & Leung, 2021), has been conducted from a qualitative standpoint. While these studies provide valuable insights, they lack the generalizability that large-scale quantitative data can offer. This is particularly problematic in the Chinese context, where the education system is often driven by high-stakes exams, leaving formative assessment practices under-researched at a broader level. The absence of large-scale, quantitative studies makes it difficult to fully understand the extent to which formative assessments are being implemented and what factors might influence their adoption. Therefore, quantitative research is urgently needed to provide a more comprehensive and representative understanding of how formative assessment is practiced in schools and how it can be effectively transferred from teachers' intention into teaching practice.

Theory of planned behavior and formative assessment

An action being taken from intention to implementation is a complex process, influenced by multiple factors. To understand this process and make connections among factors to human behavior, Ajzen (1991) proposed the Theory of Planned Behavior (TPB) model to explain the prediction of human behavior (see Figure 1). Taking music teaching as an example, this model suggests that music teachers' intentions and subsequent teaching actions are highly influenced by three core components: their personal attitude (i.e., attitude), perceived social pressure (i.e., subjective norm), and their individual belief in their abilities or confidence (i.e., perceived behavioral control). These components interact to shape music teachers' intentions. Their attitude and subjective norm will

directly influence their teaching intentions, and their perceived behavioral control will both directly and indirectly impact their teaching intentions through its influence on musical teaching behavior.

Previously, the TPB model has been applied successfully to interpret various behaviors across different contexts (Oluka et al., 2014), including formative assessment in Hong Kong (Yan & Cheng, 2015). Yan et al. (2022) further differentiated the key factors that shape teachers' intentions and practices based on personal, school, and course contexts. Despite the original three factors in TPB, they identified high-stakes accountability assessment, instructional environment, school policy and support, and student characteristics as the key contextual factors influencing Hong Kong teachers' assessment intention and implementation. However, more research is needed to verify the relevance of the additional factors. First, although Hong Kong is part of China, its education system differs significantly from Mainland China's, particularly in teaching and assessment approaches (Berry, 2011). Thus, given that formative assessment has also been promoted in the Chinese Mainland for decades, a contextual exploration is necessary to understand mainland teachers' formative assessment practices. Second, Yan et al. (2022) used a small convenience sample ($N=296$) with a limited representation of arts-related subjects. Music teachers' intentions and practices regarding formative assessment remain largely unexplored. To address this, the exploration needs to extend to the music teaching field. Thus, this study employs an extended version of TPB, covering both personal factors and contextual factors, as the theoretical framework.

Purpose of the study

This study aims to use the TPB model to investigate teachers' intentions regarding the implementation of formative assessment in Mainland Chinese schools' music teaching contexts. Additionally, since the mediating role of intention clarifies how influencing factors translate into practices and reveals gaps between intentions and actions, this study also examines the mediating effects of formative assessment intention between influencing factors and actual practices in Mainland Chinese music education. Figure 2 presents the model with three hypotheses:

H1: Music teachers' intention to apply formative assessment is predicted by their personal factors (i.e., instrumental attitude, subjective norm, self-efficacy) and contextual factors (i.e., high-stakes accountability assessment, instructional environment, school policy and support, and student characteristics).

H2: Music teachers' formative assessment practice is predicted by their intentions, self-efficacy, and all contextual factors.

H3: Music teachers' formative assessment intention mediates the relationships between formative assessment practice and each predictor of formative assessment.

Materials and methods

Participants

A total of 671 music teachers from 29 cities/provinces responded to the survey. Geographically, the survey covered North and Northeast China (46), East China (106), Central and South China (430), Southwest China (67), and Northwest China (22). The sample comprised 443 (66.02%) elementary and 228 (33.98%) secondary school teachers. The numbers of female and male teachers were 589 (87.78%) and 82 (12.22%), respectively. Their teaching experience ranged from 1–5 years (161,

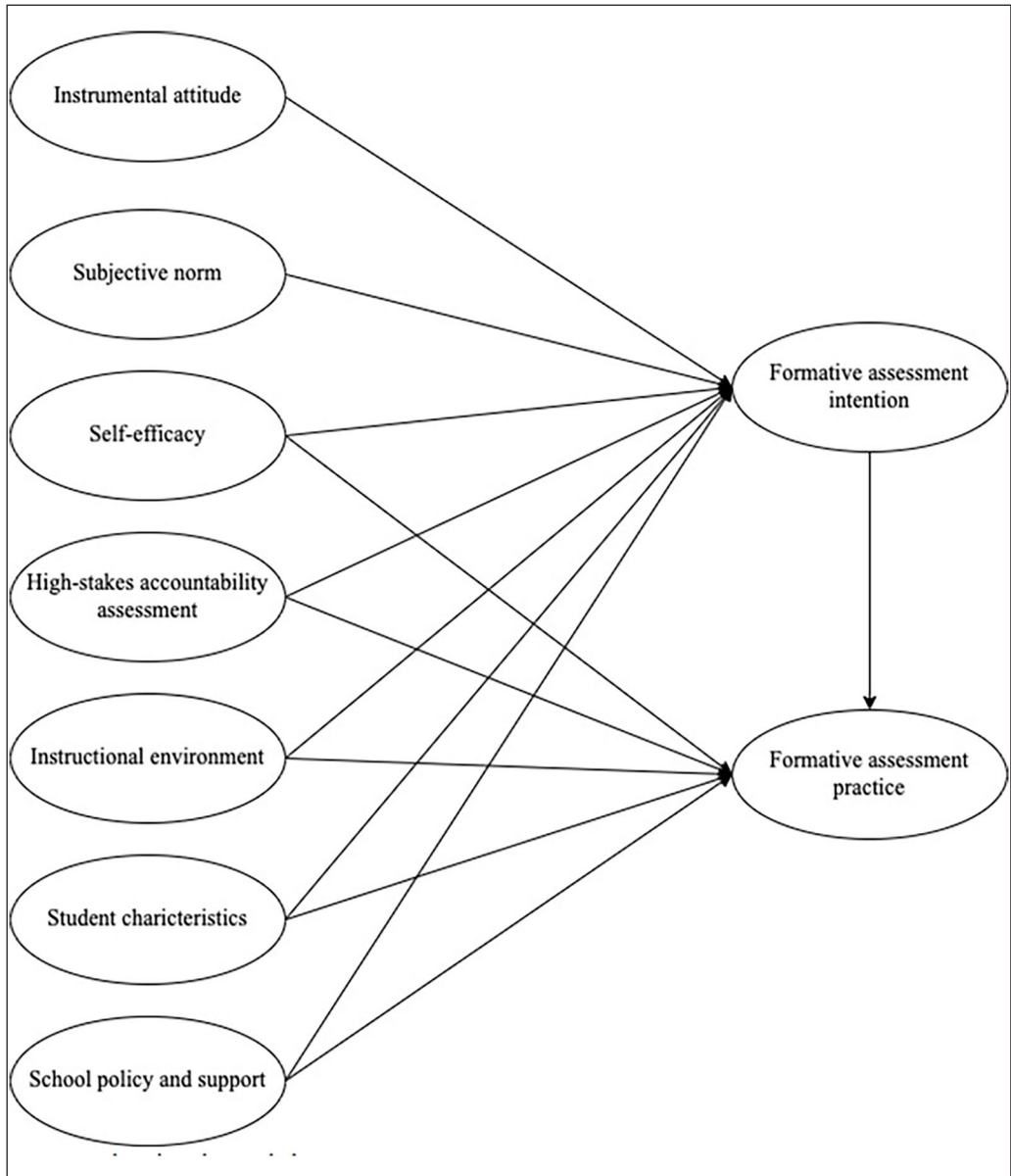


Figure 2. Hypothesized model.

23.99%) to 6–10 years (133, 19.82%), 11–15 years (117, 17.44%), 16–20 years (101, 15.05%), and above 20 years (159, 23.70%).

Instrument

The measurement instrument applied in this study contained three parts. First, four predictors of formative assessment were assessed using the Teacher's Conceptions and Practices of Formative

Assessment Questionnaire (Yan & Cheng, 2015): (a) instrumental attitude (10 items; Rasch reliability=0.88; e.g., “Formative assessment encourages students to work harder”), (b) self-efficacy (6 items; Rasch reliability=0.84; e.g., “I have enough time to implement formative assessment”), (c) subjective norm (5 items; Rasch reliability=0.75; e.g., “My colleagues support the implementation of formative assessment”), and (d) intention (6 items; Rasch reliability=0.88; e.g., “I am willing to make an effort to implement formative assessment”). The responses were measured on a Likert scale from 1 (strongly disagree) to 6 (strongly agree).

Second, to evaluate how often teachers engage in formative assessment practices, the Teacher Formative Assessment Practice Scale (Yan & Pastore, 2022) was employed. This concise, theory-based scale consists of 10 items (Cronbach’s $\alpha=0.77$; e.g., “I ensure homework can reflect students’ learning progress”) and encompasses the five key strategies outlined in Wiliam and Thompson’s (2008) formative assessment framework. The responses are recorded on a Likert scale ranging from 1 (never) to 6 (very frequently).

Third, considering the Chinese Mainland’s high-stakes examination culture and top-down educational system, four scales developed by Yan et al. (2022) were used to assess contextual factors. These scales cover examination culture (5 items; Cronbach’s $\alpha=0.79$; e.g., “Students care more about the final examination result instead of using formative assessment to improve their learning”), school support (5 items; Cronbach’s $\alpha=0.86$; e.g., “School management teams support the implementation of formative assessment”), teaching environment (5 items; Cronbach’s $\alpha=0.83$; e.g., “Before class, I have enough time to prepare for implementing formative assessment”), and student attributes (5 items; Cronbach’s $\alpha=0.84$; e.g., “My students positively participate in my class formative assessment activities”) (Yan et al., 2022). This Likert scale ranges from strongly disagree (1) to strongly agree (6).

The above instruments were originally designed for the context of Hong Kong. Given the differences between the educational systems of Hong Kong and the Chinese Mainland, two rounds of revision were conducted to ensure the validity of the content of this instrument. The first round was revised by a university-based music expert conducting music education and assessment research, who ensured that the writing style matched the Chinese Mainland context. The second round was checked by two experienced mainland music teachers with over 15 years of teaching experience, who ensured the readability of the instruments for Chinese school music teachers.

After these revisions, some wording adjustments were made (e.g., “Education Bureau Curriculum Guide” was changed to “National Curriculum Standard”), and three items were newly added. These comprised two items measuring the instructional environment (i.e., “My inspectors’ support provides me with the opportunity to implement formative assessment” and “Demonstration lesson support provides me with the opportunity to implement formative assessment”) and one for subjective norm (i.e., “My inspectors support the implementation of formative assessment”).

The revised instrument included the following components: formative assessment intention (FAI), formative assessment practice (FAP), High-stakes accountability assessment (HSAA), instructional environment (IE), school policy and support (SPS), student characteristics (SC), instrumental attitude (IA), self-efficacy (SE), and subjective norm (SN), which contained 6, 10, 5, 7, 5, 5, 10, 6, and 6 items, respectively. Items collecting demographic information, such as gender, teaching experience, teaching grades, and previous formative assessment training experience, were also included at the beginning of the instrument.

Procedure

Data were collected through a Chinese online questionnaire survey platform, Wenjuanxing, in 2023. Ethical approval was sought and given by the first two authors’ affiliated universities. The survey was sent out through a public online application, WeChat Public Platform. Informed

consent was obtained, and the participants were informed that the data collected would be used only for research purposes, that they had the right to withdraw from the study at any time without any negative consequences, and that no identifiable information would be disclosed.

Data analysis

Before analyzing the data, we measured skewness (sk) and kurtosis (ku) to assess the normality of the data distribution using the *moment's* package (Komsta & Novomestky, 2015). Kim (2013) suggested that data can be considered non-normally distributed when the absolute values of sk and ku are larger than 2 and 7, respectively. Subsequently, two analytical methods, that is, confirmatory factor analysis (CFA) and structural equation modeling (SEM), were implemented in the *lavaan* package (Rosseel, 2012) to answer the hypotheses.

The following two-step approach (Anderson & Gerbing, 1988) was used: (1) the measurement properties of the measurement model were examined via CFA and (2) SEM was used to examine the structural relations among the constructs (see Figure 1). Because the results of the normality test indicated a violation of the normal distribution (i.e., four items' ku values were larger than 7, see Appendix A. Table B1), maximum likelihood estimation with robust standard errors (MLR) was used for CFA and SEM (Abdullah et al., 2022). Multiple fit indices were used to examine model fit, including the chi-square by degrees-of-freedom value (required to be smaller than 3), the Tucker–Lewis index (TLI, required to be larger than .90), the comparative fit index (CFI, required to be larger than .90), the standardized root mean square residual (SRMR, required to be smaller than 0.08), and the root mean square error of approximation (RMSEA, required to be smaller than 0.08) (Hu & Bentler, 1999; McDonald & Ho, 2002; Wang & Wang, 2019). In addition, composite reliability (CR) and average variance extracted (AVE) were measured to assess the convergent power (Fornell & Larcker, 1981). Cronbach's α coefficient evaluated internal consistency. Discriminant validity was assessed by comparing the square root of AVE with the correlation coefficients between dimensions. Finally, a bootstrap test with 5,000 samples was performed to test hypothesis 3 (Hayes, 2009). All the calculations were conducted in RStudio version 2023.09.0+463 (Posit Software, PBC).

Results

Psychometric properties of the measurement model

Before hypothesis testing, CFA was conducted to test the psychometric properties of the nine constructs: IA, SN, SE, HSAA, IE, SC, SPS, FAI, and FAP. The measurement model fitted the data well, $\chi^2=4069.640$, $df=1611$, $\chi^2/df=2.526$, $RMSEA=.059$, $CFI=.911$, $TLI=.905$, $SRMR=.077$. The factor loadings of all items were higher than .400 (.506–.967) (see Appendix A. Table A1). As shown in Table 1, the Cronbach's α coefficients for the constructs ranged from .828 to .981, indicating satisfactory internal consistency. Additionally, the CR values for the constructs fell within the range of .837 to .977, surpassing the threshold value recommended by Fornell and Larcker (1981) and providing evidence of construct reliability. Moreover, each value of AVE also exceeded the .360 threshold (Fornell & Larcker, 1981), indicating that the convergent validity was acceptable. The square root of each AVE was higher than its corresponding correlation coefficients, showing good discriminant validity (see Table 1).

Descriptive statistics and correlations

Table 2 provides each construct's means, standard deviations, and intercorrelation coefficients. The mean score values ranged from 4.203 (SPS) to 5.170 (FAI). For the predictors, teachers exhibited

Table 1. Reliability values for the main variables in the study.

Factors	Cronbach's α	CR	AVE
Instrumental attitude	.978	0.977	0.807
Subjective norm	.891	0.887	0.613
Self-efficacy	.921	0.926	0.679
High-stakes accountability assessment	.828	0.843	0.523
Instructional environment	.894	0.901	0.573
School policy and support	.957	0.957	0.815
Student characteristics	.934	0.939	0.756
Formative assessment intention	.981	0.837	0.462
Formative assessment practice	.908	0.902	0.489

Note. $n=760$; CR = composite reliability; AVE = average variance extracted.

the lowest level of agreement with SPS (4.203) and the highest level of agreement with IA (4.887). All variables were significantly and positively correlated.

Model testing

Having examined the validity of the measurement models, we performed SEM (see Figure 3) to test our hypotheses. The results demonstrated sufficient model data fit: $\chi^2=4233.629$, $df=1614$, $\chi^2/df=2.526$, RMSEA=.061, CFI=.905, TLI=.899, SRMR=.077. In addition, as shown in Figure 2, FAI had a positive correlation with IA ($\beta=.416$, $p<.001$), SN ($\beta=.181$, $p<.01$), and SE ($\beta=.290$, $p<.001$), but a negative correlation with SP ($\beta=-.317$, $p<.001$). Moreover, FAP had significant positive relationships with SE ($\beta=.460$, $p<.001$), SPS ($\beta=.227$, $p<.001$), and FAI ($\beta=.108$, $p<.05$). There was no significant difference between the model with control for demographic factors (i.e., gender, teaching grade, curriculum standards training, and formative assessment training) and that without control ($\chi^2=4.254$, $df=3540$, $p=1.000$).

Interestingly, for the relationship between SPS and FAI, the path coefficient in SEM (negative) and the correlation coefficient (positive) did not have the same sign. A possible reason is that the original relationship between the two had been suppressed. As Falk and Miller (1992) suggested, this relationship was examined, and the discrepancy was found to be caused by “real suppression,” that is, when the necessary predictor is eliminated, a specification error occurs (the value of R^2 changed from .361 to .332). In this case, the correct sign is indicated by the path coefficient. Therefore, SPS had a negative correlation with FAI.

Mediation testing

A significant mediating role of FAI was found for three paths (see Table 3). Specifically, via FAI, 1) IA was positively related to FAP (estimate=.050, $SE=.020$, $p=.014$, 95% CI=[.012, .092]); 2) SE was positively related to FAP (estimate=.045, $SE=.019$, $p=.018$, 95% CI=[.010, .086]); 3) SPS was negatively related to FAP (estimate=-.024, $SE=.011$, $p=.024$, 95% CI=[-.047, -.005]).

Discussion

This study applied an extended TPB model (Yan et al., 2022) to investigate whether (1) Chinese Mainland music teachers' intention to use formative assessment can be predicted by personal and contextual factors, (2) their practices can be predicted by their intentions, self-efficacy, and contextual

Table 2. Descriptive statistics and correlations for the main variables in the study.

Factors	Mean	SD	IA	SN	SE	HAS	IE	SPS	SC	FAI	FAP
Instrumental attitude	4.887	0.673	.898								
Subjective norm	4.725	0.789	.601***	.783							
Self-efficacy	4.564	0.872	.426***	.442***	.824						
High-stakes accountability assessment	4.301	0.927	.180***	.192***	.251***	.723					
Instructional environment	4.404	0.811	.517***	.558***	.619***	.263***	.757				
School policy and support	4.203	1.116	.413***	.596***	.508***	.196***	.693***	.903			
Student characteristics	4.368	0.950	.542***	.542***	.589***	.181***	.739***	.746***	.869		
Formative assessment intention	5.170	0.687	.512***	.330***	.368***	.110***	.293***	.107***	.222***	.680	
Formative assessment practice	4.216	0.842	.316***	.301***	.536***	.125***	.422***	.416***	.429***	.273***	.699

Note. IA = instrumental attitude, SN = subjective norm, SE = self-efficacy, HAS = high-stakes accountability assessment, IE = instructional environment, SPS = school policy and support, SC = student characteristics, FAI = formative assessment intention, FAP = formative assessment practice.
 ***p < .001.

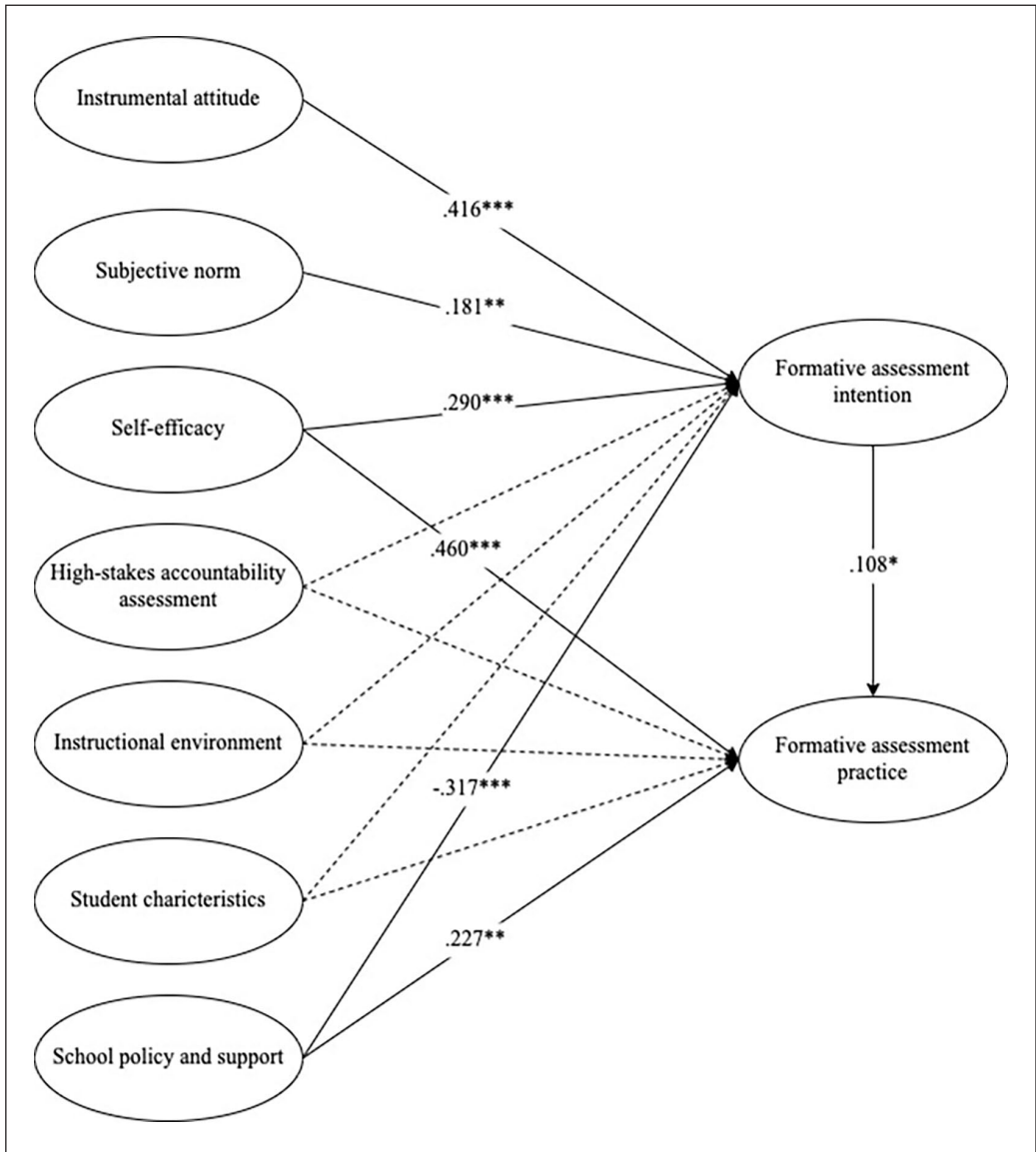


Figure 3. Structural equation modeling for the hypothesized model. Standardized path estimates are reported. Significant estimated values are shown in bold lines, and non-significant values are shown in dotted lines.

factors, and (3) intention mediates the relationship between formative assessment practice and its predictors. Unlike prior research focusing solely on the benefits of formative assessment (Parkes & Burrack, 2020; Wong, 2014), This study provides a structural understanding of the factors influencing teachers’ intention to implement formative assessment and their actual implementation in the Chinese school music teaching context. The first two hypotheses were supported, showing that the components of the extended TPB explained 37.7% of the variance in teachers’ intentions and 36.1% of that in their

Table 3. Bootstrap analyses of the magnitude and statistical significance of the indirect paths.

Path	Estimate	SE	P	95% CI
IA→FAI→FAP	0.050	0.020	0.014	[0.012, 0.092]
SN→FAI→FAP	0.021	0.011	0.055	[0.004, 0.046]
SE→FAI→FAP	0.045	0.019	0.018	[0.010, 0.086]
SPS→FAI→FAP	-0.024	0.011	0.024	[-0.047, -0.005]

Note. IA = instrumental attitude; SN = subjective norm; SE = self-efficacy; SPS = school policy and support; FAI = formative assessment intention; FAP = formative assessment practice.

practices. Highlighting the model's robustness, these findings align with Armitage and Conner (2001), who found that TPB components accounted for 39% of the variance in intention and 27% of that in behavior based on a meta-analysis of 185 studies. The results are discussed in detail below.

Personal factors

In general, music teachers with positive attitudes (i.e., instrumental attitude), supportive social surroundings (i.e., subjective norms), and high confidence (i.e., self-efficacy) are more likely to intend to engage in formative assessment. High self-efficacy also directly leads to its actual practice.

Regarding formative assessment intention, instrumental attitude was the strongest predictor ($\beta = .416$), followed by self-efficacy ($\beta = .290$) and subjective norm ($\beta = .181$). These findings are consistent with Armitage and Conner's (2001) meta-analysis of 185 studies on formative assessment intention using the TPB, but partially differed from those of Yan et al. (2022). In Yan et al.'s study, self-efficacy was the strongest predictor of teachers' intention to practice formative assessment, meaning confidence in implementing formative assessment was the primary influence. Conversely, our study found that instrumental attitude was the most influential factor, suggesting that music teachers who recognize the value and function of formative assessment are most likely to intend to apply it. This discrepancy may stem from our focus on music teachers, whereas Yan et al.'s study encompassed teachers from diverse backgrounds. Specifically, Music learning naturally involves continuous feedback from both oneself and others (Parkes & Burrack, 2020). This constant exchange of input aligns well with formative assessment, which supports ongoing skill refinement and growth. By using formative assessment, music teachers can offer real-time feedback that aids students in refining their musical skills and deepening understanding. Additionally, formative assessment allows teachers to identify individual learning needs and adjust their instruction accordingly. Given this close alignment between the feedback-driven nature of music learning and the goals of formative assessment, music teachers are more inclined to see it as an essential tool, driving their interest and willingness to implement it in their teaching practice.

Additionally, the results revealed the mediating role of intention between instrumental attitude and actual formative assessment implementation, suggesting that the stronger teachers' recognition of formative assessment, the higher chance they will use it in the classroom. Besides, as other studies reported (Karaman & Şahin, 2017; Myyry et al., 2022), our results also revealed that with higher self-efficacy in applying formative assessment, teachers are more likely to engage in formative assessment practices. Although this relationship was less well explored in music education previously, the significant impact of self-efficacy on music performance is well-documented (McPherson & McCormick, 2006). Our study strengthens the argument that teachers' self-efficacy regarding formative assessment is a critical pivot point for the uptake of formative assessment practices in classrooms.

Contextual factors

Generally, school policy and support (SPS) was the only significant predictor of formative assessment intention and implementation. However, it shows an intriguing contrast effect, with a negative effect ($\beta = -0.317$) on music teachers' intention but a positive one ($\beta = 0.227$) on formative assessment practice. This means that when schools supportively encourage implementing formative assessment, music teachers will follow it even with their opposite intentions of actually not wanting to do it. The mediation analysis further confirmed this prediction, showing that higher levels of perceived SPS led to lower music teachers' intention to practice formative assessment. Interestingly, previous research found that teachers in supportive schools showed higher formative assessment intention and practice (Brink & Bartz, 2017; Yan et al., 2022), whereas the music teachers in this study exhibited the opposite trend. The context of Chinese Mainland music teaching may explain this.

Detailly, Zhang & Leung (2023, 2024) reported that the Chinese music educational system is hierarchical and top-down, emphasizing standardized learning over individual student progress. Chinese music teachers believe they are implementing Western-based student-centered education (SCE) but operate within a collective-oriented and content-driven environment. Given that this top-down system emphasizes collective SCE, focusing on standardized outcomes for the entire class rather than individual students, teachers face evaluation pressures from national curricula (macro-level), regional inspectors (meso-level), and school-level directives (micro-level). However, as the first two levels are mandatory under the Ministry of Education, the micro-level within school instruction presents the only opportunity for teachers to make their own choices and translate their real intentions into classroom practice. Consequently, there is a high possibility that Chinese teachers might initially express negative attitudes toward new changes recommended and supported by the school. Over time, however, they may adapt to these changes and actually take action with school support, such as the formative assessment practices in this study, as they build resilience to these constraints (Yang & Zhang, 2023).

Implications for teaching and teacher professional development

This study revealed that music teachers' attitude toward the application of formative assessment was the strongest predictor of teachers' intention to implement, followed by their self-efficacy (i.e., confidence), which was the strongest predictor of actual practice. These findings have practical implications for teaching and professional development.

First, enhancing music teachers' confidence in formative assessment should be a priority, as it influences both intention and implementation. Since mastery experiences can enhance self-efficacy (Bandura & Wessels, 1997), professional development programs should provide on-site support to help teachers effectively implement formative assessment and build confidence in its use.

Second, music teachers need to develop a comprehensive understanding of the purpose of formative assessment and its practical application to increase their attitude toward using it in practice. This understanding should go beyond what is written in official documents (Ministry of Education, 2022) and be based on evidence-based outcomes and success stories. Practical training specific to assessing musical skills is also essential, as it equips teachers with necessary assessment skills and strengthens their belief in formative assessment's benefits.

Third, since this study found that teachers might resist implementing formative assessment even with support, we might need to consider the contextual constraints. Zhang & Leung (2023, 2024) found that Chinese music teachers face challenges like large class sizes, time pressure, and content-driven textbooks, which hinder adaptation to any educational reforms. This suggests that school administrators need to notice these practical challenges when making reform decisions. Since assessment practices are contagious among teachers (Yan & King, 2023), a more positive

assessment environment among teachers, students, and school culture is desirable. It is suggested that schools might first create a supportive environment with peer collaboration and a self-assessment culture for teachers to enhance their belief in formative assessment. Once a supportive, reflective, and collaborative school culture is established, teachers would most likely transfer their beliefs into practice, increasing the likelihood of effective formative assessment implementation.

Limitations

This study's limitations include its explanatory variables and the measurement of formative assessment practice. First, the explanatory variables (e.g., teachers' years of teaching, teaching grades, or past professional development) were not included in the model. Future research should consider including more explanatory variables to achieve a more comprehensive understanding of the context. Second, this study relied on teachers' self-reported data, which only captured their current thoughts without elucidating the reasons behind them or verifying whether these thoughts were consistent over time. Therefore, future research could employ qualitative methods, such as interviews, to gain deeper insights into teachers' intentions regarding and implementation of formative assessment, and record observations to verify the consistency between teachers' reported intentions and actual practices.

Conclusion

This study contributes to the literature by extending research on formative assessment into an understudied area, namely music education in the Mainland Chinese context. It offers a unique understanding of Chinese Mainland music teachers' intentions and implementation regarding formative assessment, along with the associated influencing factors. Based on an extended TPB model, the study reveals that a positive attitude toward the benefits of formative assessment, a supportive and collaborative social environment that encourages formative assessment, and strong confidence in using formative assessment enhance teachers' intention to use it in the Chinese Mainland music education context. Greater confidence in using formative assessment also directly increases its actual implementation in teaching. Given that increased school support did not significantly affect teachers' intention to implement and implementation of formative assessment practices, school administrators might consider providing more assessment training and fostering a collaborative and supportive formative assessment culture. By doing that, teachers' pressures might be eased, allowing them time to adapt to educational changes when introducing new assessment approaches.

Author contribution(s)

Le-Xuan Zhang: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Validation; Visualization; Writing – original draft.

Zi Yan: Conceptualization; Methodology; Resources; Supervision; Validation; Writing – review & editing.

Xiang Wang: Formal analysis; Methodology; Resources; Software; Validation; Writing – review & editing.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research study was supported by the Start-up Research Grant (number RG 19/2023-2024R) from The Education University of Hong Kong.

ORCID iD

Le-Xuan Zhang  <https://orcid.org/0000-0002-4394-7734>

Zi Yan  <https://orcid.org/0000-0001-9305-884X>

References

- Abdullah, H., Ibrahim, M., Harun, H., Baharun, H., Ishak, M., & Mohd Ali, S. (2022). A structural equation modeling (SEM) investigation of the L2 learning model of motivational development among Tahfiz students. *Issues in Language Studies*, 11(2), 1–19. <https://doi.org/10.33736/ils.4359.2022>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4), 471–499. <https://doi.org/10.1348/014466601164939>
- Bandura, A., & Wessels, S. (1997). *Self-efficacy*. Cambridge University Press.
- Berry, R. (2011). Educational assessment in Chinese Mainland, Hong Kong and Taiwan. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 49–61). Springer.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Booth, N., & Kinsella, V. (2022). The importance of threshold concepts and formative assessment in lower-secondary school group composing. *British Journal of Music Education*, 39(2), 145–156. <https://doi.org/10.1017/S0265051722000067>
- Brink, M., & Bartz, D. E. (2017). Effective use of formative assessment by high school teachers. *Practical Assessment, Research and Evaluation*, 22, Article 8. <https://doi.org/10.7275/p86s-zc41>
- Brown, G. T. L. (2019). Is assessment for learning really assessment? *Frontiers in Education*, 4, Article 64. <https://doi.org/10.3389/educ.2019.00064>
- Deng, Z. (2021, September 10–12). *Exploring effectiveness of formative assessment for music students in China* [Conference session]. 2021 6th International Conference on Modern Management and Education Technology (MMET 2021), Hulunbuir, China (Vol. 582, pp. 68–73). Atlantis Press. <https://doi.org/10.2991/assehr.k.211011.013>
- Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling*. University of Akron Press.
- Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18(3), 382. <https://doi.org/10.1177/002224378101800313>
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408–420. <https://doi.org/10.1080/0363775090331036>
- 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>
- Karaman, P., & Şahin, Ç. (2017). Adaptation of teachers' conceptions and practices of formative assessment scale into Turkish culture and a structural equation modeling. *International Electronic Journal of Elementary Education*, 10(2), 185–194. <https://www.iejee.com/index.php/IEJEE/article/view/320>
- Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry and Endodontics*, 38(1), 52–54. <https://doi.org/10.5395/rde.2013.38.1.52>
- Komsta, L., & Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis, and related tests. R package version 14(1). <https://cran.ma.imperial.ac.uk/web/packages/moments/moments.pdf>
- Kordeš, U., Kafol, B. S., & Brunauer, A. H. (2014). A model of formative assessment in music education. *Athens Journal of Education*, 1(4), 295–308. <https://doi.org/10.30958/aje.1-4-2>

- Lu, Y. (1987). Jiaoyu Pingjia Yingtuchu Xingchengxing Pingjia [Educational evaluation should emphasize formative evaluation]. *Tianjin Education*, 5, 6–7. <https://kns.cnki.net/kcms2/article/abstract?v=GqSshYi5sNf9NAlsUbOugpod7iiVjnT1k4AMf9LTTFNVGA9e9e5yLlZ9dGIxlht4-wAyV0cL9sUYXgIEsfpy0YyEzkoetSzyEr6myhshl4GmtUkshTAhv0KZ6x-dmhRObgsYKa-5T6AVNbNaMG7o0qEXFF0w81xD8aPfDo7E-pgpoLGMWd7cupKlsILoTrYD35ON3qtztcJcGUoXG-hN5DOzCVVmYyNNKD4Badk8OLbWvbAO0rgZBIYmXBY9IWff&uniplatform=NZKPT&language=CHS>
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64–82. <https://doi.org/10.1037//1082-989x.7.1.64>
- McPherson, G. E., & McCormick, J. (2006). Self-efficacy and music performance. *Psychology of Music*, 34, 325–339. <https://doi.org/10.1177/0305735606064841>
- Ministry of Education, PRC. (2001). *Music curriculum standards of compulsory schooling (Trial)*. Beijing Normal University Press.
- Ministry of Education, PRC. (2022). *Arts curriculum standards for compulsory education (2022 ed.)*. Beijing Normal University Press.
- Myrsky, L., Karaharju-Suvanto, T., Virtala, A.-M., Raekallio, M., Salminen, O., Vesalainen, M., & Nevgi, A. (2022). How self-efficacy beliefs are related to assessment practices: A study of experienced university teachers. *Assessment & Evaluation in Higher Education*, 47(1), 155–168. <https://doi.org/10.1080/02602938.2021.1887812>
- Oluka, O. C., Nie, S., & Sun, Y. (2014). Quality assessment of TPB-based questionnaires: A systematic review. *PLoS ONE*, 9(4), e94419. <https://doi.org/10.1371/journal.pone.0094419>
- Parkes, K. A., & Burrack, F. (2020). *Developing and applying assessments in the music classroom*. Routledge.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Wang, J., & Wang, X. (2019). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.
- William, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (1st ed., pp. 53–82). Routledge. <https://doi.org/10.4324/9781315086545>
- Wong, M. W. Y. (2014). Assessment for learning, a decade on: Self-reported assessment practices of secondary school music teachers in Hong Kong. *International Journal of Music Education*, 32(1), 70–83. <https://doi.org/10.1177/0255761413491056>
- Yan, Z., & Cheng, C. K. (2015). Primary Teachers' Attitudes, Intentions and Practices Regarding Formative Assessment. *Teaching and Teacher Education*, 45, 128–136. <https://doi.org/10.1016/j.tate.2014.10.002>
- Yan, Z., Chiu, M. M., & Cheng, E. C. K. (2022). Predicting teachers' formative assessment practices: Teacher personal and contextual factors. *Teaching and Teacher Education*, 114, 103718. <https://doi.org/10.1016/j.tate.2022.103718>
- Yan, Z., & King, R. B. (2023). Assessment is contagious: The social contagion of formative assessment practices and self-efficacy among teachers. *Assessment in Education: Principles, Policy & Practice*, 30(2), 130–150. <https://doi.org/10.1080/0969594X.2023.2198676>
- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 228–260. <https://doi.org/10.1080/0969594X.2021.1884042>
- Yan, Z., & Pastore, S. (2022). Assessing teachers' strategies in formative assessment: The Teacher Formative Assessment Practice Scale. *Journal of Psychoeducational Assessment*, 40(5), 592–604. <https://doi.org/10.1177/07342829221075121>
- Yang, Y., & Zhang, L. X. (2023). Building professional resilience: School music teachers' instructional practice development under curriculum reform. *Research Studies in Music Education*, 1321103X231209692 <http://doi.org/10.1177/1321103X231209692>
- Yuan, S., & Shu, D. (2017). Woguo Waiyu Jiaoxuezhong de Xingchengxing Pingjia Yanjiu [Research on formative evaluation in foreign language teaching in China]. *Foreign Language Teaching Theory and Practice*, 4, 51–56+21. https://kns.cnki.net/kcms2/article/abstract?v=GqSshYi5sNdTd-YSK_1AEWdZDDyXV1oZD6xD-N78cXIxivdB_qn4eaElaag5j22WnW86HH2P6XTqAC3AWujVXLkY-

IUz_ge-Ww-8-aiXIoCsbdyRiDiSXeXl86n3iyP2_KLT8A21LD73uM1vG1wwk_XLyHYQOejAS-II70
 MWpIPlhgpZC4JKwH0dllyLkWiXMgGLtAV8Ddx6jRsVxB-1anRwmZ21AHUy&uniplatform=NZK
 PT&language=CHS

- Yuan, Z., & Leung, B. W. (2021). Perceptions of developing creativity in piano performance and pedagogy: An interview study from the Chinese perspective. *Research Studies in Music Education, 45*(1), 141–156. <https://doi.org/10.1177/1321103x211033473>
- Zhang, L. X., & Leung, B. W. (2023). Context matters: Adaptation of student-centred education in China school music classrooms. *Music Education Research, 25*(4), 418–434. <https://doi.org/10.1080/14613808.2023.2230587>
- Zhang, L. X., & Leung, B. W. (2024). Defining music demonstration lessons: a unique performance-based lesson type improving teachers' instructional skills in Chinese mainland education. *British Journal of Music Education, 1*–15. <http://doi.org/10.1017/S0265051724000214>
- Zhang, L. X., Leung, B. W., & Yang, Y. (2023). From theory to practice: Student-centered pedagogical implementation in primary music demonstration lessons in Guangdong, China. *International Journal of Music Education, 41*(2), 271–287. <https://doi.org/10.1177/02557614221107170>

Appendix A

Table A1. Results of the factor loading, standardized error, z-value, and p value in the measurement model.

Instrumental attitude=~	Factor loading	SE	Z-value	p Value
IA1	0.857	0.030	21.275	.000
IA2	0.824	0.031	20.168	.000
IA3	0.832	0.031	20.312	.000
IA4	0.855	0.031	20.217	.000
IA5	0.931	0.028	24.120	.000
IA6	0.945	0.030	23.598	.000
IA7	0.948	0.030	23.796	.000
IA8	0.946	0.028	24.067	.000
IA9	0.905	0.033	19.757	.000
IA10	0.927	0.028	24.202	.000
Subjective normal =~				
SN1	0.799	0.045	14.621	.000
SN2	0.881	0.047	18.815	.000
SN3	0.841	0.045	15.343	.000
SN4	0.704	0.050	14.862	.000
SN5	0.669	0.044	14.837	.000
Self-efficacy =~				
SE1	0.650	0.041	12.049	.000
SE2	0.856	0.039	25.689	.000
SE3	0.850	0.041	18.701	.000
SE4	0.883	0.039	22.646	.000
SE5	0.817	0.048	21.149	.000
SE6	0.864	0.041	21.682	.000

(Continued)

Table A1. (Continued)

Instrumental attitude=	Factor loading	SE	Z-value	p Value
High-stakes accountability assessment =				
HSA1	0.761	0.053	16.993	.000
HSA2	0.789	0.054	17.325	.000
HSA3	0.754	0.053	15.637	.000
HSA4	0.506	0.060	12.246	.000
HSA5	0.768	0.054	15.076	.000
Instructional environment =				
IE1	0.562	0.042	10.446	.000
IE2	0.617	0.042	13.164	.000
IE3	0.582	0.038	12.779	.000
IE4	0.799	0.041	20.965	.000
IE5	0.879	0.038	24.877	.000
IE6	0.928	0.033	30.190	.000
IE7	0.839	0.037	27.640	.000
School police and support =				
SPS1	0.866	0.042	24.008	.000
SPS2	0.919	0.039	28.843	.000
SPS3	0.906	0.041	26.744	.000
SPS4	0.954	0.036	33.456	.000
SPS5	0.866	0.048	20.896	.000
Student characteristic =				
SC1	0.756	0.046	15.738	.000
SC2	0.869	0.040	20.809	.000
SC3	0.928	0.036	27.103	.000
SC4	0.929	0.035	28.212	.000
SC5	0.853	0.038	28.230	.000
Formative assessment intention =				
FAI1	0.928	0.036	19.265	.000
FAI2	0.956	0.035	19.518	.000
FAI3	0.939	0.035	19.489	.000
FAI4	0.941	0.035	18.654	.000
FAI5	0.967	0.034	20.724	.000
FAI6	0.954	0.034	19.616	.000
Formative assessment practice =				
FAP1	0.521	0.052	13.348	.000
FAP2	0.609	0.051	16.415	.000
FAP3	0.591	0.050	13.122	.000
FAP4	0.639	0.047	16.647	.000
FAP5	0.544	0.046	10.645	.000
FAP6	0.612	0.040	12.772	.000
FAP7	0.794	0.043	20.062	.000
FAP8	0.844	0.038	23.949	.000
FAP9	0.855	0.043	22.465	.000
FAP10	0.864	0.040	25.432	.000

Note. IA=instrumental attitude; SN=subjective normal; SE=self-efficacy; HAS=high-stakes and accountability assessment; IE=instructional environment; SPS=school policy and support; SC=student characteristics; FAI=formative assessment intention; FAP=formative assessment practice.

Table B1. The results of skewness and kurtosis.

Item	Kurtosis	Skewness
FAI1	6.873	-1.247
FAI2	7.116	-1.182
FAI3	6.367	-1.096
FAI4	6.997	-1.107
FAI5	6.738	-1.155
FAI6	7.022	-1.127
FAP1	2.537	-0.691
FAP2	2.283	-0.495
FAP3	4.062	-1.034
FAP4	2.714	-0.562
FAP5	6.071	-1.373
FAP6	5.611	-1.127
FAP7	3.558	-0.753
FAP8	3.471	-0.749
FAP9	3.224	-0.707
FAP10	3.162	-0.725
HSA1	3.612	-0.954
HSA2	3.343	-0.881
HSA3	3.989	-1.006
HSA4	1.990	-0.282
HSA5	4.124	-1.015
IA1	4.276	-0.609
IA2	3.911	-0.569
IA3	5.107	-0.773
IA4	4.676	-0.713
IA5	4.951	-0.740
IA6	5.456	-0.870
IA7	5.274	-0.776
IA8	5.230	-0.746
IA9	4.536	-0.623
IA10	4.531	-0.672
IE1	5.601	-1.077
IE2	4.840	-1.045
IE3	5.631	-1.147
IE4	3.805	-0.941
IE5	3.313	-0.770
IE6	3.084	-0.674
IE7	2.576	-0.587
SC1	5.270	-1.233
SC2	4.170	-0.895
SC3	3.589	-0.809
SC4	3.595	-0.849
SC5	2.768	-0.678
SE1	5.092	-0.858
SE2	3.372	-0.900
SE3	4.523	-0.935

(Continued)

Table B1. (Continued)

Item	Kurtosis	Skewness
SE4	4.491	-1.014
SE5	3.287	-0.873
SE6	4.147	-0.968
SN1	7.366	-1.402
SN2	5.658	-1.347
SN3	7.167	-1.339
SN4	4.105	-0.926
SN5	4.730	-1.050
SN6	6.087	-1.330
SPS1	3.360	-0.895
SPS2	3.101	-0.803
SPS3	3.380	-0.945
SPS4	2.998	-0.801
SPS5	4.075	-1.102

Note. IA = instrumental attitude; SN = subjective normal; SE = self-efficacy; HAS = high-stakes and accountability assessment; IE = instructional environment; SPS = school policy and support; SC = student characteristics; FAI = formative assessment intention; FAP = formative assessment practice.