International Journal of Research in Education Humanities and Commerce

Volume 06, Issue 05 "September - October 2025"

ISSN 2583-0333

INFORMAL FORMATIVE ASSESSMENT IN SINGING TEACHING: A QUALITATIVE STUDY IN MACAO

TCHIANG VAN MAN ISABEL, QI ZIXIANG & SERRA SOFIA

a Asst. Prof. Dr. TCHIANG, Van Man Isabel, School of Education, University of Saint Joseph, Macao S.A.R.
b Phd Candidate, Mr. QI Zixiang, School of Education, University of Saint Joseph, Macao S.A.R.
c Principal Investigator, Dr. SERRA Sofia, University of Aveiro, Portugal.

https://doi.org/10.37602/IJREHC.2025.6503

ABSTRACT

This study explored informal formative assessment in singing teaching in Macao. Singing activities in Macao are conducted in three types of music classes: general music classes, choral classes, and solfege classes. This study applied structured observation, and ten music classes from middle one to three were observed through recordings. The results showed that: a) the frequency and type of learning targets and success criteria varied significantly across different types of classes; b) questioning and informal performance assessment were both applied, with the latter used exclusively for assessing singing performance; c) teachers provided all feedback verbally rather than in written form. Immediate feedback was given much more frequently than delayed feedback, and directive feedback was given much more frequently than facilitative feedback; d) strengths of teacher practice included the specificity of feedback and mode of delivery (e.g., demonstration through verbal or non-verbal means), while weaknesses related to questioning strategies, specificity and level of feedback, and peer assessment. By using observation, this study compensated for the inherent limitations of self-reported data (e.g., questionnaires and interviews) in previous studies, more objectively and completely revealing the practice of informal formative assessment in a specific sphere of music education.

Keywords: Informal, performance assessment, singing, observation

1.0 INTRODUCTION

The effectiveness of formative assessment on students' learning achievement has been well-documented (Black & Wiliam, 1998a; Hattie & Timperly, 2007; Shute, 2008). Additionally, effective strategies for formative assessment have been explored within the field of music education (e.g., Green & Hale, 2009; Scott, 2012; Denis, 2018; Gallo, 2019; Martin, 2020; McPherson, 2022). The concept of formative assessment was introduced to teachers through the Music Guideline (2017) and was later stipulated as "a main method of assessment" in the Student Assessment System for Formal Education of the Local Education System (2020). These documents highlight the increasing importance of formative assessment in the curriculum of Macao. However, the actual implementation of formative assessment in singing teaching has largely been left to the discretion of teachers. Consequently, how music teachers in Macao implement formative assessment in real-time singing classes remains unknown, as do the strengths and weaknesses of its implementation. Therefore, there is a need to conduct research to explore the practice of formative assessment in singing teaching by Macao teachers. Based on the findings, further suggestions can be made to improve the current practice of formative assessment. Moreover, comparing the implementation of formative assessment in

singing teaching with studies conducted in other educational contexts may reveal important similarities and discrepancies.

2.0 LITERATURE REVIEW

2.1 Formative Assessment in Music Performance Teaching

Black and Wiliam (2009) defined formative assessment as: "to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited" (p. 25). Based on this definition, Wiliam and Thompson (2008) developed a theoretical framework for formative assessment, which includes five strategies: a) Clarifying and sharing learning intentions and success criteria, b) Eliciting evidence of students' learning, c) Providing feedback that moves learners forward, d) Activating learners as owners of their own learning, and e) Activating learners as learning resources for each other.

In terms of effectiveness, Black and Wiliam (1998) conducted a meta-analysis of 250 studies on the impact of formative assessment practices and found a positive influence on student achievement, with effect sizes ranging from 0.40 to 0.70. They argued that formative assessment interventions are more important than other educational interventions for improving student learning. This evidence has widely supported the application of formative assessment in many fields of education, including music education.

In music performance teaching, formative assessment is often implemented informally, which can be understood as short-cycle formative assessment (Wiliam, 2018, p. 51). The cycle length of such formative assessment ranges from minute-by-minute to day-to-day. Therefore, the impact of this type of assessment is less about generating data and more about helping teachers respond in real time to their students' learning needs. More importantly, as indicated by Wiliam (2018), short-cycle formative assessment should be a priority for schools and teachers, as its impact on students is greater than that of long-cycle and medium-cycle formative assessments. In music classes, it can be challenging to regularly administer summative assessments and feedback due to time and resource constraints in large ensembles. Therefore, synchronous formative assessment is often the type of assessment that students experience most in performance-based music classes.

2.2 Effective Strategies of Formative Assessment in Music Performance Teaching

Although Leahy and Wiliam (2014) systematically constructed the framework for formative assessment, its application can vary considerably from one domain to the next (Hodgen & Marshall, 2005). Therefore, effective strategies for formative assessment specifically in music performance teaching require further investigation.

Firstly, in the literature on formative assessment, teacher demonstration (both good and bad) is considered the most appropriate method for clarifying success criteria in subjects such as music and PE (Clarke, 2014, p. 132). This method has two advantages: it can entertain students and help them recognize the key features or components of certain techniques or skills (Hattie & Clarke, 2019, p. 67). Additionally, showing excellent or varied examples of art is a good way

Volume 06, Issue 05 "September - October 2025"

ISSN 2583-0333

to co-construct success criteria, as analyzing previous examples can help students extend their understanding and build their expertise (Hattie & Clarke, 2019, p. 57). Another method, referred to as eavesdropping, is considered an efficient way of gathering success criteria since students have more opportunities to articulate their thinking during the process (Clarke, 2014, p. 134). Consequently, interesting ideas can be shared with the whole class, and misconceptions can be noted so that the direction of the classes can be re-oriented accordingly.

Two methods can be used to collect evidence of learning in singing classes. In other academic subjects, the most commonly used method is questioning. Questions are also used in music performance teaching, although to a relatively lesser extent. Teachers have formulated different types of questions in music classes, such as closed questions, guided questions, and open questions (Allsup & Baxter, 2004). Questions can be categorized as: open questions that allow for broad and varied responses, guided questions that lead students to information or an answer, and closed questions that seek specific answers. Based on this categorization, the authors suggest that music teachers should structure their teaching by beginning with open questions, followed by guided questions, and then ending with closed questions. This sequence leads students through a thinking process of taking in information and then focusing that information to look for specific answers to the initial questions of wonderment. In fact, many teachers often gravitate toward closed questions, which often have only one answer. However, as recommended by Bernard and Abramo (2019), teachers are encouraged to frame more open and guided questions that allow for multiple answers.

On the other hand, performance assessment is the only assessment method that strongly matches skill targets (Chappuis & Stiggins, 2020, p. 111). Since vocal instruction encompasses many skill targets, performance assessment is considered the most important assessment method. When applying performance assessment in singing classes, many authors have indicated several advantages of using assessment tools such as checklists, rating scales, and rubrics (Green & Hale, 2009; Scott, 2012; Pellegrino, 2015; Gallo, 2019).

In relation to giving feedback (strategy three), many authors have highlighted the detrimental effects of evaluative feedback such as praise (Butler, 1987), external rewards (McPherson et al., 2012), and grades (Shaw, 2018, p. 168). Furthermore, grades should not be mixed with comments. As Butler (1988) found, students who received both grades and comments actually performed worse in the third session than those who received only grades or comments, although the difference could be small. In contrast, teachers should provide descriptive feedback to students because such feedback can answer three important questions in formative assessment: “Where is the learner going?” (Feed Up), “Where is the learner now?” (Feed Back), and “Where to next?” (Feed Forward) (Hattie & Timperley, 2007). Of these three types of feedback, Feed Forward can have some of the strongest effects on learning. Consequently, McPherson et al. (2022) emphasized that teachers should avoid repetitive comments that merely identify errors in performance and instead give students a clear indication of “where to go next” to improve the effectiveness of teacher feedback (p. 3). Nevertheless, Shaw (2018, p. 63) pointed out that more teachers provided too few Next Steps than too many. In other words, while current teacher feedback can point out flaws or weaknesses in student performance, inadequate Feed Forward still leaves students to determine their own Next Steps.

The ways in which teachers provide feedback, such as timing strategies and complexity, have also been explored in previous literature. Both delayed and immediate feedback can be given, as well as directive and facilitative feedback, and the application should vary according to the learning needs of the students. For example, immediate feedback is more appropriate for beginners who are starting to learn a new technique (Shute, 2008). Such feedback is preferable as it can ensure that students do not encode incorrect information or habits (Martin, 2019). In contrast, delayed feedback is suitable for advanced learners as it can encourage self-regulated learning and facilitate learning transfer.

In terms of complexity, feedback can be divided into directive and facilitative feedback. Directive feedback can be given in the form of verification (e.g., yes-no), correct response, error flagging, and trying again, as categorized by Shute (2008). Such feedback tells the student what to correct or revise. In contrast, facilitative feedback involves offering commentary that guides students to make independent observations and decisions (Martin, 2020). Such feedback can be given in the form of hints, cues, and prompts (Shute, 2008). In terms of effectiveness, the complexity of the feedback had no impact on student learning (Shute, 2008). Simple and focused feedback (more directive feedback) can have a greater effect on student learning achievement than complex feedback (more facilitative feedback).

Regarding activating learners as owners of their own learning (strategy four), a previous study confirmed that criteria-referenced self-assessment had a small positive effect on students' musical achievement (Valle, 2015; Valle et al., 2016). Andrea (2010) explained three steps of self-assessment: articulating and understanding expectations, critiquing one's own performance in terms of learning expectations, and using self-generated feedback as a guide for revision. Specifically in music subjects, students are encouraged to record themselves and then listen to the recordings while following the score (Denis, 2018). Additionally, teachers have been advised to provide students with effective self-assessment tools, such as roadmaps or rating scales (Shaw, 2018). Similar to self-assessment, peer assessment (strategy five) also had a small positive effect on students' musical achievement (Valle, 2015; Valle et al., 2016).

Andrea (2010) explained three steps of self-assessment: articulating and understanding expectations, critiquing one's own performance in terms of learning expectations, and using self-generated feedback as a guide for revision. Specifically in music subjects, students are encouraged to record themselves and then listen to the recordings while following the score (Denis, 2018). Additionally, teachers have been advised to provide students with effective self-assessment tools, such as roadmaps or rating scales (Shaw, 2018).

Similar to self-assessment, peer assessment (strategy five) also had a small positive effect on students' musical achievement (Valle, 2015; Valle et al., 2016; Hsia, 2016). Its implementation modalities are also similar to self-assessment. Peer assessment has been included in the Model Cornerstone Assessments (MCAs). For example, in the second grade MCA, students are required to provide verbal feedback to help their peers improve. Self-assessments in the form of practice logs are also included in the Ensemble MCAs at all grade levels, with peer assessment interjected to guide improvements before the final summative assessment.

Effective strategies for peer assessment include modeling and discussing effective and ineffective peer feedback (Leahy & Wiliam, 2015), providing students with structured

protocols (Leahy & Wiliam, 2015; Shaw, 2018), and offering sentence starters for peer feedback (Leahy & Wiliam, 2015).

2.3 Singing Teaching and Formative Assessment in the Educational Contexts of Macao

Singing is regarded as an important teaching component in Macao's music curriculum. The Requirements of Basic Academic Attainments (2015) pointed out the learning targets of singing in junior middle school. First, students should be able to sing expressively with clear diction, accurate intonation, rhythm, phrasing, and using good singing techniques (DSEDJ, 2015). Additionally, the document emphasized that students should participate in two-part choral singing. While singing choral repertoire, students should be aware of dynamics and timbre in multiple voice parts, and cooperate accurately with the conductor and piano accompanist. Based on this, this study investigates multiple types of music classes, including general music classes, solfege classes, and choral classes.

The concept of formative assessment was first introduced to Macao in 2017 after the Music Guideline (2017) was released by DSEDJ. This document regulated formative assessment as a salient assessment type in contrast to summative assessment. Even though this document defined formative assessment, indicated its advantages, and suggested multiple methods of applying it, how formative assessment should be implemented according to different school contexts was still left for teachers to consider. Tchiang (2017) explored teachers' perspectives on these education reforms, highlighting their views on the challenges and opportunities presented by the new assessment policies. In the latest regulation called "Student Assessment System for Formal Education of Local Education System," formative assessment was defined as "a type of continuous assessment which is carried out constantly in the course of learning and teaching and focuses on the process of learning." More importantly, this regulation also stipulated that student assessment shall combine formative assessment with summative assessment, with the former being the primary type of assessment (DSEDJ, 2020). Tchiang (2019) further emphasized the need for continuous professional development and support to effectively integrate formative assessment into teaching practices.

The 2024 revision of the Basic Academic Attainment Requirements (BAA) for music introduced several changes compared to the 2016 version. The updated BAA places a stronger emphasis on student-centered learning and the integration of technology in music education. It also provides more detailed guidelines on the use of formative assessment to support student learning and development. Additionally, the 2024 version includes updated criteria for evaluating student performance, with a focus on creativity and critical thinking skills. (DSEDJ, 2024)

Research Question

RQ 1 How do teachers implement informal formative assessment (FA) in multiple types of music classes?

RQ 2 What are the strengths and weakness of implementing informal FA by teachers?

3.0 METHODOLOGY

The study adopted content analysis methods (Krippendorff, 2018) to examine the implementation of informal formative assessment in music teaching. Elo and Kyngas (2008) defined content analysis as "a research method for making replicable and valid inferences from data to their context, with the purpose of providing knowledge, new insights, a representation of facts and a practical guide to action." It contains a quantitative feature in which the frequency of data can be identified and systematically collected, and can be interpreted by using an inductive coding system and deductive testing system of the qualitative method (Briggs et al., 2012). As the study aims to investigate the implementation of informal formative assessment in real classes, content analysis through observational data was conducted. Such a method is regarded as a direct, reflective, and effective measuring tool that allows observers to capture the varying elements of the teaching process and describe people's behavior and the context through many categorized lenses (Cakir, 2010).

In this study, ten classes were recorded by the second author with the consent of the teachers who taught these classes. The specific method of data collection was non-participant observation (Creswell, 2016). The researcher acted outside of the demonstration setting. Data collection was conducted by observing the video recordings without direct interaction with the classroom activities or the teachers. In this way, any external interruption of the observation could be avoided in order to maintain the original pedagogical status (Creswell, 2016). A total of 10 video recordings of music classes, each with a duration of 40 minutes, were analyzed. They were categorized into three different types of classes: four general music classes, three solfege classes, and three choral classes (Table 1).

Table 1. Description of the observed class

Teacher	Class	Class topic	Class Type	Duration (Min)	Instructional language
Ms. A	Class A	Breathing in singing	General Music Class	40	Chinese-Mandarin
Ms. B	Class B	*"Wo Ai Ni Zhong Guo"*	Choral repertoire rehearsal	40	Chinese-Cantonese
Ms. C	Class C	*"Qi Dai De Yuan Fang"*	Choral repertoire rehearsal	40	Chinese-Mandarin
Ms. D	Class D	*"Tai Yang Chu Lai Xi Yang Yang"*	Choral repertoire rehearsal	40	Chinese-Mandarin
Ms. E	Class E	5/8 meter	Solfeggio in choral teaching	40	Chinese-Mandarin
Ms. F	Class F	*"O Shi Fen Zhong"*	General Music Class	40	Chinese-Cantonese
Ms. G	Class G	*"Qing Chun Wu Qu"* (part one)	General Music Class	40	Chinese-Mandarin
	Class H	*"Qing Chun Wu Qu"* (part two)	General Music Class	40	Chinese-Mandarin
Mr. H	Class I	Major Scale	Solfeggio in choral teaching	40	Chinese-Cantonese
Ms. I	Class J	*"Da Hai A Gu Xiang"* and *"Liang Zhu"*	General Music Class	40	Chinese-Cantonese

Volume 06, Issue 05 "September - October 2025"

ISSN 2583-0333

During data analysis, 10 video recordings were imported into Nvivo 14 to code and categorically analyze classroom behaviors. In the current study, a quantitative sequential analysis was conducted (Bakeman & Quera, 2011). This method was applied to particular observational data to quantify behavior. Firstly, a coding scheme was developed based on previous literature (Bernard & Abramo, 2019; Hattie & Timperley, 2007; Martin, 2019; Chappius & Stiggins, 2020). The specific codes for each category were derived from the data during the analysis process (Table 2).

Table 2. Specific Codes for Data Analysis

Category	Codes
Clarifying learning intentions and success criteria	Knowledge learning target Skill learning target Disposition learning target Success criteria related to technical accuracy Success criteria related to expressive qualities No criteria or criteria not specified
Eliciting learning evidence	Open questions Close questions Guided questions Informal performance assessment
Providing feedback that movers learner forward	Praise at personal level Praise related to task Feed Up at task level Feed Up at process level Feed Back at task level Feed Back at process level Feed Forward at task level Feed For ward at process level Teacher demonstration Immediate feedback Delayed feedback Directive feedback Facilitative feedback
Activating learners as the resources of each other	Peer assessment

To ensure the trustworthiness of the study, a school music teacher with more than 10 years of teaching experience was invited to participate in an "interrater agreement" by applying Cohen's Kappa's point-to-point agreement (Bakeman & Quera, 2011) with the second author regarding the codes of Feed Up at task level, Feed Up at process level, Feed Back at task level, Feed Back at process level, Feed Forward at task level, and Feed Forward at process level. In this process, 50 items (25.9%) of feedback delivered in these ten observed classes were selected by the second author. Codes were displayed in an agreement matrix, tallying both agreement and disagreement, and calculated using Cohen's Kappa's agreement statistic (Bakeman & Gottman, 1997). The mean values of the six codes obtained overall agreement rates of 83.5%, 93.4%, 89.1%, 71.9%, 100%, and 74.4% for Feed Up at task level, Feed Up at process level, Feed Back at task level, Feed Back at process level, Feed Forward at task level, and Feed

Forward at process level, respectively. These values indicated good reliability between the two raters.

4.0 RESULTS

RQ 1 How do teachers implement informal FA in multiple types of music classes?

The learning objectives of the observed classes varied depending on the type of teaching. In general music classes, skill or technique was not the only goal. These classes also included other learning objectives such as knowledge objectives or dispositional objectives. For example, in classes G and H, the song "Qing Chun Wu Qu" was sung, while the main objective was to let students experience the beauty of the Xin Jiang folk song or express their creativity through music. The mastery of singing techniques was therefore not particularly emphasized in these classes. In contrast, choral classes were mainly concerned with presenting a certain repertoire fully and expressively. Therefore, vocal skills and techniques were emphasized in these classes to achieve an ideal musical quality. In solfeggio classes, the teaching objectives focused on the application of musical skills and music theory knowledge. Thus, students were expected not only to understand musical concepts such as 5/8 time (class E), half step, major scale, and minor second (class I), but also to apply these concepts through performance tasks such as conducting, reading, writing, or singing set by the teachers.

Based on various learning objectives, teachers applied success criteria related to both technical accuracy and expressive qualities. As shown in Table 3, teachers most frequently applied success criteria related to technical accuracy 114 times (55.6%). Among the success criteria related to technical accuracy, teachers most frequently applied pitch accuracy (36%), followed by phonation methods (e.g., breathing, resonance, and singing posture) (30%), and rhythmic accuracy (26%). Diction was also used by the teachers, albeit with a lower frequency (8%). Additionally, teachers less frequently applied success criteria related to expressive qualities 65 times (31.7%). Specifically, teachers most frequently applied dynamics (47%), followed by timbre (20%) and phrasing (14%). Less frequently, teachers applied success criteria related to articulation (6%) and musical emotion (13%). Furthermore, no criteria were used 19 times (9.3%) and unspecified criteria were used 7 times (3.4%).

Table 3. Total Amount of Applied Success Criteria by Classes (times)

	Success criteria related to Technical Accuracy						Success criteria related to Expressive qualities					No criteria	Criteria not specified
	pitch	rhythm	diction	Phonation method			phrasing	dynamics	timbre	articulation	Musical emotion		
				breathing	resonance	Singing posture							
Class A	0	0	0	7	1	0	0	1	0	0	0	2	1
Class B	1	1	1	1	1	0	6	13	6	0	5	4	1
Class C	10	5	0	0	0	0	1	3	0	0	0	2	1

Class D	3	1	4	9	0	2	2	9	4	0	2	0	2
Class E	0	11	0	0	0	0	0	0	0	0	0	2	0
Class F	0	1	0	1	0	1	0	1	0	2	0	3	0
Class G	8	4	5	1	2	1	0	1	0	0	0	0	1
Class H	0	2	0	0	0	1	0	0	0	2	1	2	0
Class I	18	4	0	1	2	0	0	1	1	0	0	4	0
Class J	1	1	0	1	0	1	0	2	2	0	0	0	1
Total	41	30	10	21	6	6	9	31	13	4	8	19	7
				114					65			19	7

As shown in Table 4, success criteria were most frequently applied in choral classes, followed by solfeggio classes and general music classes. More importantly, the frequency of success criteria related to technical accuracy or expressive qualities varied among the different class types. In the choral classes, since the students had mastered the repertoire to a certain extent, they were mostly able to sing with accurate intonation and rhythm despite occasional mistakes. Therefore, teachers more often pointed out problems related to the expressive qualities of a particular repertoire rather than correcting technical errors. In contrast, solfeggio classes focused on technical accuracy. As a result, teachers applied significantly more success criteria related to technical accuracy than those related to expressive qualities. Students had to understand the theory of these concepts (e.g., the 5/8 meter in class E and the major scale in class I) and, more importantly, apply them to various tasks such as singing, reading, beating time, or conducting. In general music classes, teachers also gave more success criteria related to technical accuracy than those related to expressive qualities.

Table 4. Frequency of success criteria by class types (times/per class)

Class Type	SC related to technical accuracy (times/per class)	SC related to expressive qualities (times/per class)	No criteria or criteria not specified (times/per class)	Total (times/per class)
Choral repertoire rehearsal	13	17	3.3	33.3
Solfeggio in choral teaching	18	1	3	22
General Music Class	7.8	2.4	2	12.2

When collecting evidence of learning, teachers used informal performance assessment to evaluate students' singing performance through various learning tasks. This assessment was usually done in short cycles, from minute to minute or second to second. Teachers assessed students' performance through their observation and judgement without using any assessment tools (e.g., rubrics or rating scales). Most of the learning tasks were related to singing, but teachers also assessed students' accuracy in conducting (Class E) and dance movements (Class I). Based on such assessment, teachers provided synchronous feedback in verbal or non-verbal

form. Regarding timing strategies, teachers provided both immediate and delayed feedback (Appendix 1). Immediate feedback was given directly after a trial and was the most frequently used type of feedback during the class. Delayed feedback, on the other hand, was given a few minutes after the entire piece of music had been played and was used significantly less frequently than immediate feedback. In terms of specificity, this feedback was more general than immediate feedback, as teachers often commented on the overall quality of the performance. In terms of the complexity of feedback, teachers provided both directive and facilitative feedback. Directive feedback included reviewing students' performance, error flagging and giving explicit instruction. On the other hand, teachers also gave facilitative feedback although such feedback was given far less frequently by teachers than directive feedback (Appendix 2).

RQ 2.1 What are the strengths of implementing informal FA by teachers?

The amount of descriptive feedback varied according different class types (Table 5). Teachers gave descriptive feedback significantly more frequently in choral rehearsals (42.6 times/per class) than in solfeggio teaching (23.5 times/per class) and general music classes (15 times/per class).

Table 5. Frequency of descriptive feedback by class types (times/per class)

Class type	Frequency of giving descriptive feedback (times/per class)
choral class	42.6
Solfeggio class	23.5
General music class	15

According to the feedback framework established by Hattie and Timperley (2007), descriptive feedback can be divided into three types: Feed Up, Feed Back, and Feed Forward. As can be seen in the observed classes (Table 6), teachers gave more Feed Up (96 times) than Feed Back (72 times) and Feed Forward (52 times). Such a pattern was also observed in general music classes and choral classes, except that in Solfeggio classes teachers gave more Feed Back (20 times) than Feed Up (16 times) and Feed Forward (11 times).

Table 6. Frequency of descriptive feedback types by class types

Class type	Feed Up (times)	Feed Back (times)	Feed Forward (times)
choral class	52	42	34
Solfeggio class	16	20	11
General music class	28	10	7
Total	96	72	52

Teacher demonstration was given as a form of non-verbal feedback that was most frequently applied in choral classes, followed by Solfeggio classes and general music classes (Table 7). As observed in the study, teacher could demonstrate through various ways. Often, teachers might demonstrate to sing one musical phrase, a few notes, or even one note with sol-fa or lyrics. Except for singing, teachers also demonstrated the skills or technique such as breathing skill (Teacher A), movements of Xin Jiang Dance (Teacher G), and conducting pattern of 5/8 (Teacher F) according to diversified teaching purposes of the classes. Also, teachers could also

demonstrate through body movements (Appendix 3). Furthermore, teacher demonstration could be divided into two types: Feed Up and Feed Back while "Feed Up" demonstration was far more delivered than "Feed Back" demonstration. For Feed Up ones, teachers always demonstrated the right way of singing for students (telling students "Where they were going"). In contrast, teachers also imitated the wrong way or weakness in students' singing which was named as "Do it wrong". Such demonstration mirrored current situation of students, therefore it could be understood as another type of Feed Back (telling students "where they were"). By delivering such demonstration, teachers intended to prevent the students from repeating the mistakes and improve the weakness in their singing so that students might more easily figure out the discrepancy between their current level between the ideal level. For example, teacher B applied two different kinds of timbre to read a word "You Shi" and then explained which timbre would be better to be applied in singing.

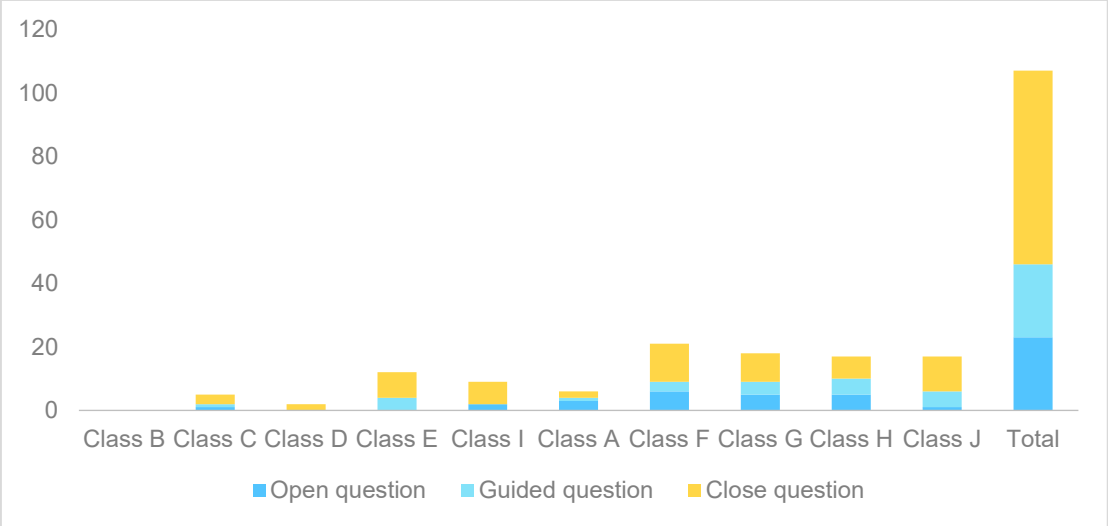
Table 7. Frequency of Teacher Demonstration by Class Types

Class Type	Demonstration (Feed Up) Times /per class	Doing it wrong (Feed Back) Times /per class	Total Times /per class
General Music Class	2.1	0	2.1
choral class	13	2.7	15.7
Solfeggio class	9.5	0.5	10
Total	24.6	3.2	27.8

RQ 2.2 What is the weakness of implementing informal FA by teachers?

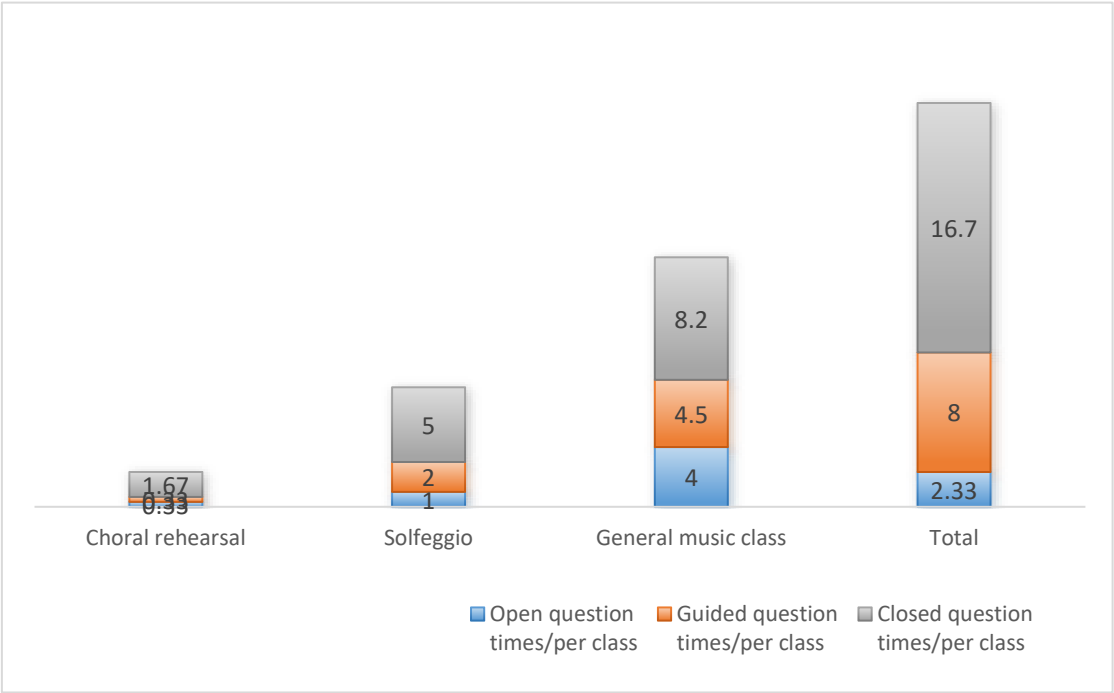
While eliciting students' understanding, the teachers asked questions. In general, teachers most frequently formulated close questions, followed by open questions and guided questions (Figure 1). Close questions seek to find unambiguous, convergent, and specific answers. They often more pointedly ask students to recall information or content, such as "What is the second note? (Class I)", "Do you think the instruments applied in this segment is string or percussion? (Class J)", and "During our reading just now, did you notice a rhythm that is quite representative. What rhythm is it? (Class F)". Guided questions steered the students in a direction that the teacher was aiming for. They pushed the students toward a particular course that hinted at information that the teacher wanted the students to notice or figure out on their own without telling them outright. For example, "What are the musical emotions, rhythmic features, and vocal style of this song? (Class G)", "Please look at the difference between bar 12 and here (bar 21)? (Class C)", and "What kind of sublimation did our music undergo when we added real percussion instruments? (Class H)". Open questions were broad and required divergent thinking and varied answers. Teachers often used such questions to assess what students already know, gather various ideas or solutions to problems, and seek students' opinions and views. For example: "Do you have any good suggestions? (Class H)", "Let us first listen to what kind of artistic conception is presented in the introduction? (Class J)", and "What do you think of first when it comes to breath? (Class A)"

Figure 1. Types of Questions by Each Class (Times)



Besides, the teachers framed different types of questions according to different types of classes (Figure 2). Among all types of classes, questions were most frequently framed in general music classes (16.7 times/per class), followed by solfeggio classes (8 times/per class) and choral classes (2.33 times/per class). In particular, close questions (8.2 times/per class) were also more frequently framed than guided questions (4.5 times/per class) and open questions (4 times/per class) in general music classes. Similarly, close questions (5 times/per class) were also more frequently framed than guided questions (2 times/per class) and open questions (1 time/per class) in solfeggio classes. Within choral classes, close questions (1.67 times/per class) was more frequently framed than guided questions (0.33 times/per class) and open questions (0.33 times/per class).

Figure2. Types of Questions by Class Types (Times/per class)



Regarding questioning by teachers, some teachers did not give students enough wait time to elaborate on the questions. In four of the classes (classes E, G, H, and J), the teachers did not ask the students to answer a question after it had been formulated. Instead, they answered the questions themselves. In this case, the students did not have time to think about the question and did not have the opportunity to answer it. For example, the teacher asked a closed question, "What marking appeared at the end of the score?" and then answered the question herself: "It is a repeat marking" (class H).

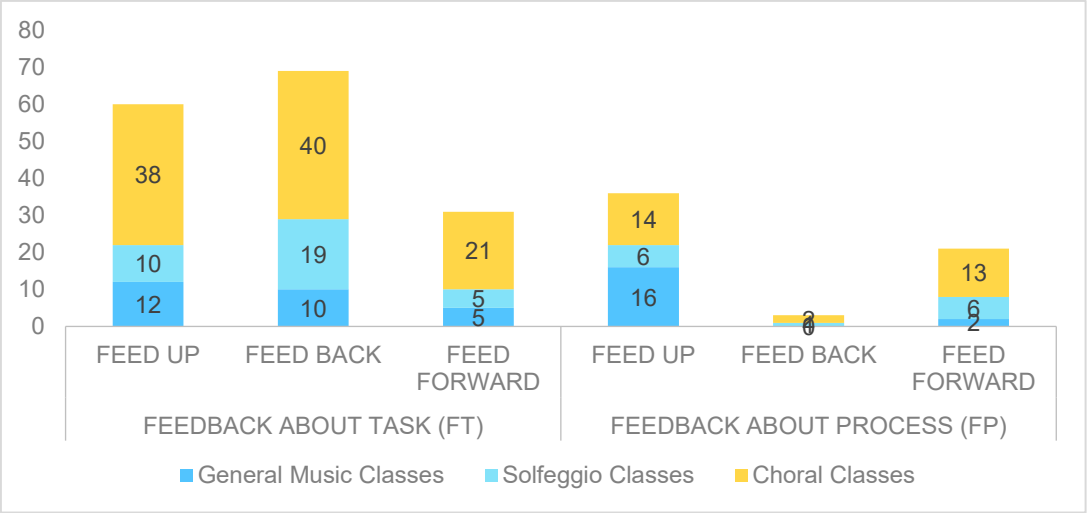
While the teachers provided a lot of descriptive feedback, they also used some vague words, albeit at a relatively low frequency (Appendix 4). These words could be categorized as "adverbs," "adjectives," and "nouns," which could confuse students or lead to different interpretations. Additionally, evaluative feedback was used by most teachers, although its frequency was significantly lower than that of descriptive feedback. The most common form of evaluative feedback was praise (Appendix 5). This praise was general and often given on a personal level, containing little information about the student's learning process. However, some praise was task-related and could contain success criteria (Appendix 6). In terms of the level of feedback, teachers in each class provided feedback about tasks (FT) and feedback about processes (FP) (Appendix 7). As shown in Table 7, teachers gave significantly more FT (160 times) than FP (60 times). Specifically, teachers provided more feedback on how well a task was completed or performed than on the strategies or students' completion of the task (Hattie & Timperley, 2007).

Table 7. Feedback Level by Each Class

	Feedback about Task (FT)			Feedback about process (FR)		
	Feed Up	Feed Back	Feed Forward	Feed Up	Feed Back	Feed Forward
Class B	9	13	11	4	0	6
Class C	12	17	7	2	0	1
Class D	17	10	3	8	2	6
Class E	7	9	3	3	1	2
Class I	3	10	2	3	0	4
Class F	2	2	0	1	0	0
Class A	1	0	0	5	0	0
Class G	2	1	2	4	0	2
Class H	5	5	3	5	0	0
Class J	2	2	0	1	0	0
Total	60	69	31	36	3	21
	160 (72.7%)			60 (27.3%)		

When combining feedback levels and types, teachers gave more Feed Back than Feed Forward at the task level for all class types (Figure 3). In particular, Feed Back was given most often in choral classes and solfeggio classes, while Feed Up was given most frequently in general music classes at this level. In contrast, at the process level teachers were more likely to give Feed Forward than Feed Back for all lesson types. Teachers at this level were also most likely to give Feed Up.

Figure 3. Feedback Types by Feedback Level



Of all the classes observed, peer assessment was only used informally in class J. This involved the teacher asking one student (Student C) to rate the singing works of her peer (Student A and B) according to a rating scale shown by the teacher (awarding points from one to five in terms of each success criteria) and make some comments on their performance. Although student C was able to provide some feedback, it was generalized as students used many judgmental adjectives such as "very good" and "very serious" without mentioning the specific musical locations of the strength. Furthermore, no weaknesses were analyzed and highlighted in the peer feedback. There were also no suggestions for further improvement in the peer feedback. Finally, the teacher rather than student C indicated the weaknesses in the students' performance and gave brief suggestions for further improvement.

5.0 DISCUSSIONS AND IMPLICATION FOR TEACHING

This study has shown that teachers in Macao use informal formative assessment in teaching singing. First, teachers used different learning objectives for different classes. In particular, skill targets are mainly used in choral classes. In solfeggio classes, skill targets were also combined with knowledge targets. The learning targets in general music classes were the most diverse, including skill targets, knowledge targets, and disposition targets. Accordingly, the success criteria were applied based on various learning targets. Teachers applied more success criteria related to technical accuracy than those related to expressive qualities. This result is consistent with the findings of Cranmore and Wilhelm (2017) and can be explained by the level of the students. Since most of the students in Macao were non-musical students who did not have a strong musical foundation, teachers had to make great efforts to promote the students' technical accuracy (e.g., pitch, rhythm, and phonation).

While tracking students' learning progress, teachers often used two methods: questioning and informal performance assessment. Regarding the type of feedback, all feedback was given verbally, and no written feedback was observed. In terms of timing strategies, teachers gave immediate feedback more often than delayed feedback. Specifically, immediate feedback was given during the teaching process after a specific musical task (Martin, 2019). This feedback included more musical details about the success criteria. Delayed feedback was given much less frequently compared to immediate feedback. Such feedback was given after students had completed a whole repertoire (Martin, 2019). Consequently, such feedback was general rather

than specific, as it could only describe the overall quality of students' performance. In terms of feedback complexity, teachers provided far more directive feedback (e.g., verification, error flagging, or explicit instruction) than facilitative feedback (e.g., prompt, clue, or hint). This result could also be explained by the limited musical foundation or cognitive abilities of the students. Novices or struggling students need support and explicit guidance during the learning process (Knoblauch & Brannon, 1981; Moreno, 2004); thus, hints may not be as helpful as explicit, directive feedback.

Strengths in teachers' practice were identified in strategy three (providing feedback that moves learners forward). Firstly, teachers gave significantly more descriptive feedback than evaluative feedback. Descriptive feedback has been shown to have a positive impact on student learning achievement (Leahy & Wiliam, 2014; Clarke, 2014; Brookhart, 2017; Hattie & Clarke, 2019; McPherson et al., 2022). Additionally, teachers frequently provided "Feed Up" through nonverbal demonstration (e.g., singing and body movements). Teacher demonstrations can be categorized into two types: "Do it right" or "Do it wrong." The first type answers an important question in formative assessment, namely, "Where is the learner going?", while the second type answers the question, "Where is the learner now?" Teacher demonstration has been widely endorsed in the literature on both formative assessment (e.g., Clarke, 2014; Hattie & Clarke, 2019) and music teaching (e.g., Sang, 1987; Dickey, 1992). Teachers should therefore maintain the strengths already mentioned and build on them at a higher level. To achieve better specificity of feedback, teachers could: a) use many nouns and descriptive adjectives instead of pronouns (e.g., "this" or "that"); b) describe concepts or criteria related to learning tasks; and c) describe learning strategies that might be useful (Brookhart, 2017). Furthermore, as Dickey (1992) suggests, musical models (both appropriate and inappropriate) need to be accurate.

Therefore, teachers need to improve their skills to correctly and incorrectly demonstrate one musical performance variable (e.g., pitch, rhythm, tempo, style) while consistently modeling all other musical performance elements.

Weaknesses in teachers' practice were identified in strategies two (eliciting learning evidence) and three (providing feedback that moves learners forward). When questioning students, teachers most frequently framed closed questions, followed by guided questions and open questions. In other words, far more questions were asked at a factual level, that is, remembering or recalling some knowledge that had been learned before. By contrast, open questions or guided questions that could help students achieve higher-order skills such as comparing, analyzing, reasoning, evaluating, and creating were less frequently applied (Holcomb, 2019). Additionally, open questions or guided questions could engineer discussion in the classes, reducing the application of the traditional "I-R-E" (Initiate-Response-Evaluation) questioning pattern (Mehan, 1979). As suggested by Leahy and Wiliam (2014), teachers could apply strategies such as "Pose-Pause-Pounce-Bounce" to get students more engaged in class discussions. Furthermore, regarding the sequence of questions, Bernard and Abramo (2019) suggested that teachers should shape questions from open, guided, to closed ones.

In terms of providing feedback, teachers were observed giving evaluative feedback at a personal level in the form of praise. Such feedback was confirmed to have no impact on students' learning performance as it contained too little information about the learning process

(Butler, 1987; Hattie & Timperley, 2007; Shute, 2008; Martin, 2019). Therefore, teachers should not underestimate the negative effects of such feedback, which could dilute meaningful feedback (Martin, 2019; McPherson et al., 2022). Praise should only be given for the purpose of developing rapport (McPherson et al., 2022). Apart from the fact that teachers mainly gave descriptive feedback, they still provided much more Feed Up and Feed Back than Feed Forward ("Next Step"). This finding is consistent with Shaw's (2018) observation that "more teachers provide too few Next Steps than too many." Students need to know exactly what to do to improve errors rather than focusing on the errors themselves (McPherson et al., 2022). Thus, teachers should include Feed Forward in their feedback (Shaw, 2018). More importantly, such feedback should be within the student's ZPD (Zone of Proximal Development), providing scaffolding rather than scolding to students who need it.

In addition, the level of teacher feedback was monotonous, mostly at the task level (FT). Feedback on process (FP) was given much less frequently than FT, and feedback on self-regulation (FR) was not observed in any classes. This phenomenon could be explained by the students' musical knowledge, as all students in the observed classes were not music students and may not have had a strong musical foundation to accept and apply the teacher's feedback at a higher level. This finding is consistent with McPherson (2022), who noted that task-level feedback is most common in learning situations. The author also pointed out that it is valuable for teachers to provide more process and self-regulation feedback as the student masters more. Too much feedback within a level can actually detract from performance (Hattie & Timperley, 2007). In summary, teachers should differentiate feedback levels to meet the learning needs of students at different stages of learning.

Regarding peer assessment, the frequency of implementing was the lowest. Also, teachers applied peer assessment as peer rating, that is, asking students to rate peer' assessment according to success criteria shown in a rating scale. Although students gave some peer feedback during the process, such feedback could be evaluative and vague. Most importantly, students did not have opportunity to give suggestions to their peers for improvement. Without such salient step, peer assessment was not complete and effective since it could not lead to any revision or improvement (Andrea, 2010; Valle et al., 2016).

6.0 CONCLUSION

This study highlights the current practices and areas for improvement in the implementation of informal formative assessment in music education within Macao. Teachers demonstrated a range of learning objectives tailored to different types of music classes, with a notable emphasis on technical accuracy over expressive qualities. This focus aligns with the needs of non-musical students who require foundational skills in pitch, rhythm, and phonation.

However, the study identified several weaknesses in teachers' practices, particularly in eliciting learning evidence and providing feedback that moves learners forward. Teachers predominantly used closed questions, limiting opportunities for students to engage in higher-order thinking and class discussions. Additionally, the feedback provided was often evaluative and lacked the specificity needed to guide student improvement effectively. The over-reliance on task-level feedback and the scarcity of process and self-regulation feedback further highlight the need for a more differentiated approach to meet students' varying learning needs.

To enhance the effectiveness of formative assessment in music education, teachers should adopt strategies that encourage open and guided questioning, provide detailed and actionable feedback, and incorporate more Feed Forward elements. Emphasizing feedback within the student's Zone of Proximal Development (ZPD) and balancing different levels of feedback can significantly improve student learning outcomes.

Overall, this study underscores the importance of continuous professional development for teachers to refine their assessment practices and better support student learning in music education. By addressing these areas, Macao's education system can foster a more holistic and effective approach to teaching music, ultimately enhancing students' musical abilities and overall educational experience.

REFERENCE

1. Allsup, R. E., & Baxter, M. (2004). Talking about music: Better questions? Better discussions! Music Educators Journal, 91(2), 29-33.
2. Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In Handbook of formative assessment (pp. 90-105). Routledge.
3. Bakeman, R., & Quera, V. (2011). Sequential analysis and observational methods for the behavioral sciences. Cambridge University Press.
4. Bakeman, R., & Gottman, J. M. (1997). Observing interaction: An introduction to sequential analysis. Cambridge university press.
5. Bernard, C. F., & Abramo, J. M. (2019). Teacher evaluation in music: A guide for music teachers in the US. Oxford University Press.
6. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: principles, policy & practice, 5(1), 7-74.
7. Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education), 21, 5-31.
8. Brookhart, S. M. (2017). How to give effective feedback to your students. Ascd.
9. Briggs, A., & Coleman, M. (2019). Research methodology in educational leadership and management. In Oxford Research Encyclopedia of Education.
10. Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. Journal of educational psychology, 79(4), 474.
11. Cakir, I. (2010). Criticizing ourselves as teachers through observation: From reflective to effective teaching. English for Specific Purposes World, 9(30), 1-12.
12. Chappuis, J., Stiggins, R. J., Chappuis, S., & Arter, J. (2020). Classroom assessment for student learning: Doing it right-using it well. Upper Saddle River, NJ: Pearson.
13. Clarke, S. (2014). Outstanding formative assessment: Culture and practice. Hachette UK.
14. Cranmore, J., & Wilhelm, R. (2017). Assessment and feedback practices of secondary music teachers: A descriptive case study. Visions of research in Music Education, 29(1), 6.

15. Creswell, J. (2016). Research in education: Design, conduct and evaluation of quantitative and qualitative research (Translated by Kouvarakou, N.). Ion (Year of Publication of the Original 2005), Athens.
16. Denis, J. M. (2018). Assessment in music: A practitioner introduction to assessing students. Update: Applications of Research in Music Education, 36(3), 20-28.
17. Dickey, M. R. (1992). A review of research on modeling in music teaching and learning. Bulletin of the Council for Research in Music Education, 27-40.
18. Direcção dos Serviços de Educação e Juventude. [DSEJ] (2015). The Requirements of Basic Academic Attainments (subject:Art Education- II.Music). http://www.dsedj.gov.mo/crdc/edu/BAA_primary/checklist/art_music_checklist_e.pdf
19. Direcção dos Serviços de Educação e Juventude. [DSEJ] (2017a). Junior Secondary School Music Curriculum Guide. Retrieved from http://www.dsej.gov.mo/crdc/guide/data/junior/Music_guide_c.pdf
20. Direcção dos Serviços de Educação e Juventude. [DSEJ] (2020). Student Assessment System for Formal Education of Local Education System. Retrieved from https://www.dsedj.gov.mo/~webdsej/www/edulaw/202007/download/Regulation_28_2020e.pdf
21. Direcção dos Serviços de Educação e Juventude. [DSEJ] (2024). Student Assessment System for Formal Education of Local Education System. Retrieved from https://portal.dsedj.gov.mo/webdsejspace/site/ddca/index.jsp?con=requirements&
22. Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. Journal of advanced nursing, 62(1), 107-115.
23. Gallo, D. J. (2019). Formative assessment practices and children's singing accuracy: A mixed methods inquiry. International Journal of Music Education, 37(4), 593-607.
24. Green, S. K., & Hale, C. L. (2011). Fostering a lifelong love of music: Instruction and assessment practices that He a difference. Music Educators Journal, 98(1), 45-50.
25. Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112.
26. Hattie, J., & Clarke, S. (2018). Visible learning: feedback. Routledge.
27. Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. Curriculum journal, 16(2), 153-176.
28. Holcomb, A. (2019). Assessment practices of American choral music educators. The Oxford Handbook of Assessment Policy and Practice in Music Education, Volume 2, 379.
29. Hsia, L. H., Huang, I., & Hwang, G. J. (2016). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses. British Journal of Educational Technology, 47(4), 618-632.
30. Knoblauch, C. H., & Brannon, L. (1981). Teacher commentary on student writing: The state of the art. Freshman English News, 10(2), 1-4.
31. Krippendorff, K. (2018). Content analysis: An introduction to its methodology. Sage publications.
32. Leahy, S., & Wiliam, D. (2009). Embedding formative assessment. London, UK: Specialist.
33. Martin, L. (2020). Using synchronous formative feedback to facilitate student growth. Music Educators Journal, 107(2), 51-57.
34. Mehan, H. (1979). 'What time is it, Denise?": Asking known information questions in classroom discourse. Theory into practice, 18(4), 285-294.

35. McPherson, G. E., Davidson, J. W., & Faulkner, R. (2012). Music in our lives: Rethinking musical ability, development and identity. Oxford University Press.
36. McPherson, G. E., Blackwell, J., & Hattie, J. (2022). Feedback in music performance teaching. Frontiers in psychology, 13, 891025.
37. Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. Instructional science, 32(1), 99-113.
38. Pellegrino, K., Conway, C. M., & Russell, J. A. (2015). Assessment in performance-based secondary music classes. Music Educators Journal, 102(1), 48-55.
39. Sang, R. C. (1987). A study of the relationship between instrumental music teachers' modeling skills and pupil performance behaviors. Bulletin of the Council for Research in Music Education, 155-159.
40. Scott, S. J. (2012). Rethinking the roles of assessment in music education. Music Educators Journal, 98(3), 31-35.
41. Shaw, B. P. (2018). Music assessment for better ensembles. Oxford University Press.
42. Shute, V. J. (2008). Focus on formative feedback. Review of educational research, 78(1), 153-189.
43. Tchiang, V. M. I. (2017). Teachers' Perspective on the Macao Education Reform Policy. PEOPLE: International Journal of Social Sciences, 3(2), 1309-1322.
44. Tchiang, V. M. I. (2019). The Challenges for Teachers on the Implementation of Education Reform Policy in Macao. Paper presented at ISER-735th International Conference on Education and Social Science (ICESS-2019), Seoul, South Korea.
45. Thompson, M., & Wiliam, D. (2007, April). Tight but loose: A conceptual framework for scaling up school reforms. In annual meeting of the American Educational Research Association. Chicago, IL.
46. Wiliam, D. (2019). Embedded formative assessment. Solution tree press.
47. Valle, C. (2015). Effects of criteria-referenced formative assessment on achievement in music. State University of New York at Albany.
48. Valle, C., Andrade, H., Palma, M., & Hefferen, J. (2016). Applications of peer assessment and self-assessment in music. Music Educators Journal, 102(4), 41-49.

