Paradigm Academic Press
Research and Advances in Education
ISSN 2788-7057
NOV. 2024 VOL.3, NO.11

Formative Assessment Practice in Singing Teaching in Secondary School: A Convergent Mixed Methods Study in Macao

Zixiang Qi[1], Serra Sofia[2] & Tchiang Van Man[1]

[1] University of Saint Joseph, Macao SAR, China

[2] University of Aveiro, Aveiro, Portugal

Correspondence: Zixiang Qi, University of Saint Joseph, Macao SAR, China.

doi:10.56397/RAE.2024.11.04

Abstract

This study is an exploration of the implementation of formative assessment in singing teaching by teachers in Macao. We applied convergent mixed methods, and the research questions focused on the current implementation of formative assessment (strengths and weaknesses). Quantitative data (N=57) was collected through questionnaires adapted from Ramsey and Duffy (2016), and qualitative data (N=9) was collected through semi-structured interviews. Both types of data were collected from in-service local music instructors who teach music in secondary schools. Subsequently, both sets of data were compared side-by-side through a matrix to address the research questions. The results suggest that: a) teachers applied all five strategies of formative assessment from Thompson and Wiliam's (2007) framework, albeit with varying frequencies; b) teachers prioritized skill targets and technical accuracy over expressive qualities; c) performance assessment dominated singing classes, while questioning was more common in other settings; d) the strengths of formative assessment implementation could be found in the clarification of learning intentions and criteria for success, and also in feedback that propelled learners forward, eliciting the learning evidence, and mobilizing students as the owners of their own learning as well as that of their peers'. Weaknesses in practice were identified, and several strategies are suggested for further improvement. Applying such strategies could increase formative assessment's effectiveness in singing lessons in Macao.

Keywords: formative assessment, secondary, singing, teaching, mixed methods

1. Introduction

Formative assessment has a profound effect on students' learning achievement (Black & Wiliam, 1998a; Hattie & Timperly, 2007; Shute, 2008). In terms of practicing formative assessment, Thompson and Wiliam (2007) proposed a systematic framework that included five key strategies. In music education field, several effective strategies of formative assessment have been investigated as well (e.g., Green & Hale, 2009; Scott, 2012; Denis, 2018; Gallo, 2019; Martin, 2020; McPherson, 2022). Within the educational context of Macao, formative assessment plays a salient role in its curriculum. Although the concept of formative assessment was introduced to teachers through *Music Guidelines (2017)*, its actual implementation in singing was left to teachers. Besides, how music teachers in Macao implemented it within singing teaching remained unknown. Consequently, there is a need to investigate the current practice of formative assessment in singing in Macao and analyze the strength and weakness in a systematic framework. Based on the findings in this research, further suggestions could be made to improve the effectiveness of implementing formative assessment in singing teaching in Macao. The implication of this study would shed light on the development of further professional training in formative assessment.

2. Literature Review

2.1 An Introduction to Formative Assessment

Black and Wiliam (2009) have framed a definition of formative assessment: "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). Wiliam and Thompson (2007) developed a theoretical framework for formative assessment that can be used to assess its implementation. This framework includes five strategies: Strategy one: clarifying and sharing learning intentions and success criteria; Strategy two: eliciting evidence of learning; Strategy three: providing feedback that can propel learners forward; Strategy four: activating students as owners of their learning; and Strategy five: activating students as instructional resources for one another. Black and Wiliam (1998) presented a meta-analysis of 250 studies on the effects of formative assessment practices. They found that it benefited students' achievement, with effect sizes ranging from .40 to .70. The authors argued that formative assessment is more important for improving student learning than for other educational interventions. Shute (2008) suggested that the typical effect size of formative feedback ranges from 0.4 to 0.8. Meanwhile, Hattie and Timperley's (2007) review of meta-analyses found an average effect size of 0.95 standard deviations across 4,157 studies. These findings have widely supported the application of formative assessment in many areas of education, including music education.

2.2 Effective Strategies of Formative Assessment in Music Performance Teaching

Teacher demonstration (both good and bad) has been regarded as the most appropriate method for clarifying success criteria in subjects such as music and P.E. (Clarke, 2014). This method has two advantages: it can "entertain students and help them to recognize the key features or ingredients of certain techniques or skills" (Hattie & Clarke, 2018, p. 67). In addition, showing exceptional or unique examples of art is a good way to co-construct success criteria because analysing previous examples can help students scaffold understanding and develop their expertise (Hattie & Clarke, 2018). Another way, eavesdropping, has been viewed as "an efficient way to gather success criteria, as 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 identified, so that a lesson's direction can be re-oriented accordingly.

While eliciting learning evidence, many authors of other subjects (e.g., English, Maths, or Science) have suggested applying effective questioning and engineering classroom discussion. However, this is not the case in singing lessons, which involve many skill targets. Indeed, performance assessment is "the only method that is a strong match for skill targets" (Chappuis & Stiggins, 2020, p. 111). Thus, performance assessment has been viewed as the main method of assessing singing performance. In the application of performance assessment in singing, many authors have pointed out the benefits of using assessment tools. As Scott (2012) has shown, "well-designed assessment tools such as rubrics can encourage students to engage with the feedback information that gives students experience in implementing actions to improve further performance" (p. 33). Furthermore, assessment tools can help students internalize success criteria for assessment (Green & Hale, 2009; Valle et al., 2016; Gallo, 2019). As recommended by Green and Hale (2009), each level of quality for each assessment criterion in assessment should be demonstrated in concrete, behavioral terms that students can understand. One of the best ways to do this is to have a discussion with students to define the level of quality.

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). Furthermore, grades should not be mixed with comments. As Butler (1988) found, students who received both grades and comments did even worse in the third session than those who received only grades or comments, although the difference could be small. In contrast, the study indicated that teachers should provide descriptive feedback to students because such feedback could answer three important questions in formative assessment: "Where is the learner going?" (feed up), "Where is the learner?" (feed back), and "Where to next?" (feed forward) (Hattie & Timperley, 2007; Brookhart, 2017). Of these three types of feedback, feed forward can have some of the most powerful effects on learning. Consequently, McPherson et al. (2022) emphasized that teachers should "avoid repetitive comments merely to identify errors in performance, and instead provide a clear indication of 'where to next?' so that teacher feedback can be more effective" (p. 3). Nevertheless, Shaw (2018) pointed out that "more teachers provided too few 'next steps' than too many" (p. 63). In other words, current teacher feedback could point out the errors or weaknesses in students' performance, but insufficient "feed forward" still left 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. In particular, both delayed and immediate feedback can be given, as well as directive and facilitative feedback, and their application should be varied according to students' learning needs. For instance, immediate feedback is more appropriate for beginners who are starting to learn a new technique (Shute, 2008). In contrast, delayed feedback is suitable for advanced learners, as it can cultivate self-regulated learning and facilitate learning transfer. In addition, the complexity of the feedback had no effect on students' learning (Shute,

2008). In fact, simple, concise, and focused feedback (more directive feedback) could have a greater effect on students' learning achievement than complex feedback (more facilitative feedback).

Previous studies have confirmed that criteria-referenced self-assessment has 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 a score (Denis, 2018). In addition, teachers have been advised to provide students with effective self-assessment tools, such as road maps or rating scales (Shaw, 2018). Similarly, criteria-referenced peer assessment also has a small positive effect on students' musical achievement (Valle, 2015; Valle et al., 2016; Hsia, 2016). Its implementation modalities are also similar to those of self-assessment. The peer assessor has to indicate both strengths and weakness in need of further improvement. After receiving the constructive peer feedback, the performer needs to illustrate what they have learned from the feedback and their further goal. Effective peer assessment strategies include modelling and discussing effective and ineffective peer feedback (Leahy & Wiliam, 2015), providing students with structured protocols (Leahy & Wiliam, 2015; Valle et al., 2016; Shaw, 2018) and providing sentence starters for peer feedback (Leahy & Wiliam, 2015).

2.3 Formative Assessment in Educational Contexts in Macao

The concept of formative assessment was first introduced into the music education field in Macao in 2017 after the DSEDJ released Music Guidelines (2017). This document regulated formative assessment as a salient assessment type, as opposed to summative assessment. Although this document defined formative assessment, outlined its benefits, and suggested several methods for its application, the decision on how to implement formative assessment in the various school contexts was left to the teachers. In the latest regulation, titled Student Assessment System for Formal Education of the Local Education System (2020), formative assessment was defined as "a type of continuous assessment that is carried out constantly over the course of learning and teaching and focuses on the learning process". This regulation also stipulated that assessment should be a combination of formative and summative assessment in Macao, with the former being the primary assessment type (DSEDJ, 2020). As can be seen from the history of formative assessment in Macao, the educational bureau has been aware of formative assessment's importance in education and has therefore published several relative documents and regulations.

3. Research Purpose and Research Questions

The goal of this study is investigating the implementation of formative assessment in singing lessons by teachers in Macao. Since this study was designed to view a single phenomenon from two different methodological perspectives, the research questions were the same in both the quantitative and the qualitative components:

RQ 1: How do teachers implement formative assessment in singing in secondary school in Macao?

RQ 2: What are the strengths and weaknesses of implementing formative assessment of teachers in singing in Macao?

4. Methodology

Mixed methods research enables the incorporation of the strengths of both qualitative and quantitative methods into a research project (Creswell, 2003; Tashakkori & Teddlie, 2017). In combining these methodologies, researchers intend to mitigate both approaches' weaknesses and view a problem from several vantage points. In this study, both the quantitative approach (which explores the frequency of formative assessment strategies applied by teachers) and the qualitative approach (which explicates how teachers apply these strategies), offered insight into this understudied phenomenon. Additionally, this mixed methods study also entailed a convergent parallel design. We applied concurrent timing to implement the quantitative and qualitative components concurrently during the research process. Additionally, we allocated each method equal weight, and kept the research strands independent during analysis, but then integrated the results for the overall interpretation (Creswell & Clark, 2007).

Quantitative data was collected through a survey using a questionnaire adapted from Ramsey and Duffy (2016). Content validity was checked by three music education experts. A pilot study (N=20) was then conducted to examine the questionnaire's reliability. After the questionnaire was revised, it was officially published on an app called Wen Juan Xing. A total of 57 questionnaires were completed and returned. According to the DSEDJ, there are 60 middle schools in Macao, and each school usually has one or two music teachers. Therefore, such a sample seemed diverse enough to represent Macau's middle school music teacher population. After developing and administering the final survey, we calculated Cronbach's alpha to measure the internal consistency of the five survey item clusters which were intended to represent various underlying constructs. The overall alpha coefficient for these sections of the survey exceeded 0.7, indicating that each cluster's reliability met the

requirements. Qualitative data were collected in parallel with the quantitative data through semi-structured interviews. Stratified purposive sampling was used to represent the experiences of important subgroups such as years of teaching experience, undergraduate major, highest degree earned, and whether participants were working as choral teachers. Finally, 9 participants were selected to participate in the interviews (Table 1). These interviews were all conducted via online videoconferencing software with the participants. Each interview lasted approximately 30-40 minutes and was recorded with participants' consent. The recordings were transcribed by the researchers and the text was presented to the participants to verify that it accurately reflected their original meaning. A traditional deductive analysis was then conducted. The qualitative data was coded sentence-by-sentence and themes were generated.

Table 1. Interview Participant Demographic Information

No	Name	Gender	Years of teaching	Undergraduate major	Highest degree earned	Teaching choir
1	Ms. A	Female	7	Music education	Bachelor's degree	Y
2	Mr. B	Male	5	Music performance (vocal)	Master's degree	Y
3	Ms. C	Female	7	Music education	Bachelor's degree	Y
4	Mr. D	Male	7	Music education	Master's degree	Y
5	Mr. E	Male	8	Music education	Master's degree	Y
6	Ms. F	Female	6	Music education	Bachelor's degree	Y
7	Ms. G	Female	13	Music performance (Piano)	Master's degree	Y
8	Ms. H	Female	9	Music administration	Master's degree	Y
9	Mr. I	Male	10	Music education	Master's degree	Y

4.1 Quantitative Data Summary

Descriptive analysis was conducted to determine the frequency at which teachers employed formative assessment strategies (Appendix 1). The results revealed that teachers applied teacher-directed formative assessment far more frequently than student-directed assessment. Teachers highly frequently applied demonstrations and showed models to articulate learning targets and criteria for success. While eliciting evidence of students' learning, both questioning and in-class performance were frequently applied. During performance assessment, teachers assessed far more success criteria related to technical accuracy, rather than expressive qualities. Moreover, they relied on general impressions more frequently than formal assessment tools. In terms of giving feedback, teachers mostly gave verbal feedback rather than written feedback. They also reported that they frequently pointed out the objects in need of improvement. Additionally, teachers both provided students with both "feed back" and "feed forward", with the latter type more frequently. Evaluative feedback was also applied by teachers, though sometimes at low frequency. Regarding delivery modes of feedback, teachers frequently gave both immediate and delayed feedback. Also, they highly frequently gave both directive and facilitative feedback.

4.2 Qualitative Findings Summary

A deductive thematic analysis was applied to the qualitative data. Seven themes emerged from this analysis: "learning goals and success criteria applied by teachers," "Performance assessment as the main method," "teacher feedback specificity," "teacher feedback delivery modes," "tracking students' responses to teacher feedback," "teachers' self-assessment practice," and "teachers' peer-assessment practice."

4.3 Mixed Method Data Analysis

The first step in mixed methods research is separate qualitative and quantitative data according to each methodology's conventions. The data in this study were analysed to address the mixed methods research questions. To enable comparison between the quantitative and qualitative data, we created a mixed methods data matrix (Table 2). For those data that are related to the same focus group category or theme, we presented the qualitative and quantitative data side-by-side for easy comparison. To describe the converged results, and also to address data-mixing, we applied convergence labels (Fitzpatrick, 2011). In those situations where the quantitative and qualitative data directly address the same phenomenon or topic and either confirm or contradict each other, we labelled the data either "confirm" or "contradict". It is possible that there are situations in which the qualitative and quantitative data regarding a phenomenon or topic may partially confirm one another, while

also in part being contradictory. In these cases, we labeled the data as having mixed convergence. In those cases, in which the quantitative and qualitative data are neither directly confirmatory or contradictory, but instead provide different perspectives on the same phenomenon or deepen the understanding of the other, we label the data as "enhance". Therefore, there are four possible researcher-designed data convergence labels: confirm, contradict, mixed, and enhance. These labels describe the convergence of the quantitative and qualitative results (Table 2).

Table 2. Table Condensed Data Convergence Matrix by Research Question

Quantitative Survey Themes	Qualitative Codes	Alignment
RQ 1: How do teachers implement formative assessment in singing in secondary school in Macao?		
Frequency of sharing learning goals and success criteria	Learning goals and success criteria shared by teachers	Enhance
Frequency of applying questioning to assess students' understanding	Performance assessment as the main method	Enhance
Frequency of applying informal in-class assessment to assess students' performance	Performance assessment as the main method	Confirm
Frequency of delivering immediate and delayed feedback	Timing strategies for teacher feedback	Enhance
Frequency of delivering directive and facilitative feedback	Teacher feedback complexity	Enhance
Frequency of implementing self-assessment	Teachers' self-assessment practice	Confirm
Frequency of implementing peer assessment	Teachers' peer-assessment practice	Confirm
RQ 2.1: What are the strengths of implementing formative assessment of teachers in singing in Macao?		
Frequency of sharing learning targets and success criteria through multiple ways	Ways teachers share learning goals and success criteria	Confirm
Frequency of verbal feedback and written feedback	Teacher feedback delivery modes	Confirm
Frequency of pointing out objects in need of improvement	Teacher feedback specificity	Confirm
Frequency of giving feed back and feed forward	Teacher feedback specificity	Mixed
Frequency of guiding learners to critique their own work	Teachers' self-assessment practice	Confirm
Frequency of monitoring the process of peer assessment	Teachers' peer-assessment practice	Confirm
RQ 2.2: What are the weaknesses of implementing formative assessment among teachers in singing in Macao?		
Frequency of applying assessment tools	Performance assessment as a main method	Confirm
Frequency of giving praise at the self level	Teacher feedback specificity	Confirm
Frequency of grading students	Teacher feedback specificity	Confirm
Frequency of guiding students to act upon teacher feedback	Tracking students' responses to teacher feedback	Enhance
Frequency of guiding students to apply self-generated data for further improvement	Teachers' self-assessment practice	Confirm

4.4 Mixed Methods Results

RQ 1: How do teachers implement formative assessment in singing in secondary school in Macao?

The quantitative results showed that teachers frequently shared learning objectives with students before they started teaching singing lessons (M=3.89, SD=0.880). In addition, learning objectives for singing lessons were

commonly related to local academic standards (M=3.77, SD=0.945). The qualitative data enhanced the quantitative findings by revealing multiple types of learning objectives. Among all types of learning objectives, the most important objectives applied were skill-related. These objectives included phonation methods, vocal skills, and polyphonic skills. Nevertheless, teachers also employed knowledge-related objectives and dispositional objectives, albeit at a much lower frequency.

The quantitative data revealed that teachers frequently asked questions to assess the entire group's singing-related knowledge (M=4.07, SD=0.863). In addition, teachers highly frequently used follow-up questions when engaging students in discourse (M=4.26, SD=0.791). The qualitative data also enhanced the quantitative results by specifying which types of questions teachers used, as well as their personal preferences. Mr. I exemplified that he asked closed-ended questions for factual material while open-ended questions to spark class discussion. While Mr. B mentioned that he preferred to apply guided questions to encourage students to contemplate the correct answers. In contrast, many teachers, including Mr. I, Ms. G and Ms. H, did not prefer the method of asking questions in singing lessons. Instead, they emphasized the importance of practicing through performance in such classes. Even when examining students' singing-related knowledge, they highlighted that performance assessment was the most appropriate assessment method. According to the quantitative data, teachers assessed learners' singing more frequently by observing and judging their in-class singing performance (M=4.14, SD=0.833) than by utilizing live and in-class singing exams (M=3.82, SD=0.966). The qualitative data confirm that performance assessment is frequently used in the classroom. Specifically, whole team assessment is mainly used in general music lessons, while group assessment is implemented by vocal part in choir lessons. As for the quantitative data, teachers significantly applied more success criteria related to technical accuracy (M=4.32, SD=0.985) as compared with those related expressive qualities (M=4.09, SD=0.931) when assessing students' singing. The findings of qualitative data corroborated the quantitative results.

The quantitative data revealed that teachers frequently delivered feedback that was both immediate (M=3.88, SD=1.13) and delayed (M=3.98, SD=1.06), but the latter's frequency was slightly higher. Qualitative data confirmed the delivery of both feedback types, while their application frequency was not directly addressed. Delayed feedback was provided following the first attempt at a song, mostly at the beginning of the class. Therefore, such feedback could be evaluative and general. By contrast, immediate feedback was mostly given during the teaching process, and therefore such feedback was more specific. Qualitative data further indicated the teachers' preferences and the factors which had influenced their decision on timing strategies including teaching content and different learning phases (Ms. H), and teachers' belief (Ms. C, Mr. I, Ms. H, and Mr. E). Regarding the complexity of teacher feedback, the quantitative data revealed that teachers highly frequently delivered both directive feedback (M=4.18, SD=0.86) and facilitative feedback (M=4.19, SD=0.81), but the latter's frequency was slightly higher. The qualitative data also confirmed the provision of both types of feedback, although the frequency of their use was not directly addressed. In addition, the qualitative data indicated teachers' differing preferences in terms of the complexity of the feedback and the reasons for it. Many teachers (e.g., Mr. I., Ms. H., Ms. G., and Ms. A) indicated that they preferred directive feedback by flagging students' errors in their performance and giving explicit verbal instruction and giving students non-verbal demonstrations. In contrast, Mr. B preferred facilitative feedback (e.g. indirect reminder, prompt, or questions) since he felt that students improved their independent thinking in this way. Other teachers (Ms. H and Ms. A) acknowledged the benefits of facilitative feedback, although they were unsure how to give their students indirect hints or prompts. In addition, the teachers believed that they should vary their feedback's complexity depending on various factors, such as restrictions on instructional time and students' cognitive abilities.

The quantitative data revealed that teachers rarely applied student self-assessment in singing lessons (M=3.25, SD=0.94). Qualitative data partly confirmed the low frequency of applying self-assessment by most of the teachers. However, one teacher (Ms. H) asserted that she frequently applied self-assessment in her classes. Additionally, qualitative data enhanced the understanding of the factors that impeded teachers from frequently applying such a strategy including restriction of instructional time (Mr. D) and limitations of students' cognitive and inner hearing abilities (Mr. F). As with self-assessment, quantitative data showed that teachers were less likely to use peer assessment in singing lessons (M=3.35, SD=1.14). The qualitative data confirmed the quantitative results and explained the reasons for most teachers' low application frequency such as students' limited ability to provide meaningful peer feedback (Ms. A), limitations of instructional time (Ms. H), and insufficient confidence and skills to implement peer assessment as perceived by teachers (Ms. A and Mr. D). Nonetheless, Mr. I reiterated the need to use peer assessment as it was regulated in the document: *Requirements of Basic Academic Attainments (2015).*

RQ 2.1: What are the strengths of implementing formative assessment for teachers in singing in Macao?

The strengths of teachers' implementation of formative assessment were mainly examined in Strategies One, Three, Four, and Five.

The convergence of the qualitative and quantitative data revealed that the teachers had applied multiple ways to clarify the success criteria to their students. The quantitative data revealed that teachers highly frequently offered demonstrations, models, or examples of singing while they articulated the success criteria (M=4.44, SD=0.73). The qualitative data further specified how teachers applied such strategies, including demonstrating right or wrong (Mr. I) and eavesdropping (Ms. C).

The convergence of the qualitative and quantitative data indicated that teachers highly frequently gave learners feedback (M=4.39, SD=0.8). Verbal feedback was given much more frequently than written feedback (M=2.68, SD=1.2). The qualitative data confirmed these results, with the exception that only one teacher (Mr. I) provided some comments in the form of written feedback. In addition, the quantitative data revealed that feed forward (e.g., pointing out a better or different way of singing that needs improvement) (M=4.14, SD=1.00) was given more frequently than feed back (e.g., telling students what they have achieved, with specific reference to whether or not they are learning) (M=3.73, SD=1.00). The qualitative data showed mixed results on this theme. At the process level, teachers gave more feed forward than feed back (e.g., "You should first adjust the inhalation position and then stabilize the larynx position before you continue singing," and "Please pay attention to the vocal placement while singing"). In contrast, at the task level, they gave more feed back than feed forward (e.g., "Your voice is pretty good and consolidate it up again," and "Your voice sounds shrivelled"). In addition, many teachers reported that they were able to point out the object receiving the feedback, such as individuals, the whole team, or a vocal part. Nevertheless, as noted in the qualitative data, teachers could use ambiguous words (e.g., some, a few, and nice) in their feedback, which reduced the feedback's specificity.

The quantitative data indicated that teachers occasionally guided their students to point out any strengths and/or weaknesses in their own singing performance (M=3.70, SD=0.90). More frequently, teachers guided the learners to indicate both the location and the dimensions of their weaknesses in their singing performance (M=3.79, SD=0.94). The qualitative data confirmed the results that teachers (e.g., Ms. A, Mr. D, and Ms. G) guided their students to identify errors or problems in singing (e.g., a certain note, or vocal technique) according to different vocal parts. Additionally, teachers reported another two strategies while they implemented self-assessment. First, teachers recorded students' singing and replayed the videos for them before guiding them in critiquing their own work. Second, two teachers (Ms. G and Mr. D) indicated that they used self-reflective questions to guide students to engage in the self-assessment process.

The quantitative data revealed that teachers occasionally monitor the peer assessment process by circulating among the pairs, giving feedback, coaching, and sequencing activities as needed (M=3.58, SD=0.99). The qualitative data enhanced these results and specified how teachers engaged students in peer assessment and guided them to frame effective peer feedback. As with self-assessment, some teachers also used reflective prompts to guide students to engage in peer assessment. Additionally, Mr. I illustrated that he instructed students to use the terms in the success criteria to help them frame descriptive peer feedback. Ms. H. mentioned another strategy, which was to combine self-assessment and peer assessment in a group (usually 4-5 students per group). She believed that in this way, students' auditory discrimination, self-regulation and cooperative learning could be improved.

RQ 2.2: What weaknesses are there in the implementation of formative assessment among singing teachers in Macao?

The weaknesses identified in formative assessment implementation were mainly examined in Strategies Two, Three, and Four. The quantitative data showed that 42% of teachers rated their students' singing based on a general impression. In addition, teachers indicated that they used assessment tools such as rating scales (35%), checklists (16%), and rubrics (7%) (Figure 1). The qualitative data confirmed that many teachers (Ms. A, Mr. B, Ms. G, and Mr. I) applied only general impressions to assess their students' performance without using any assessment tools, although some of them acknowledged some subjectivity. In addition, some other teachers mentioned that they also used assessment tools when assessing students' performance. The qualitative data enhanced the quantitative data by shedding light on the purposes for which teachers use such tools. As reported by most of the teachers, they applied assessment tools "summatively" rather than formatively.

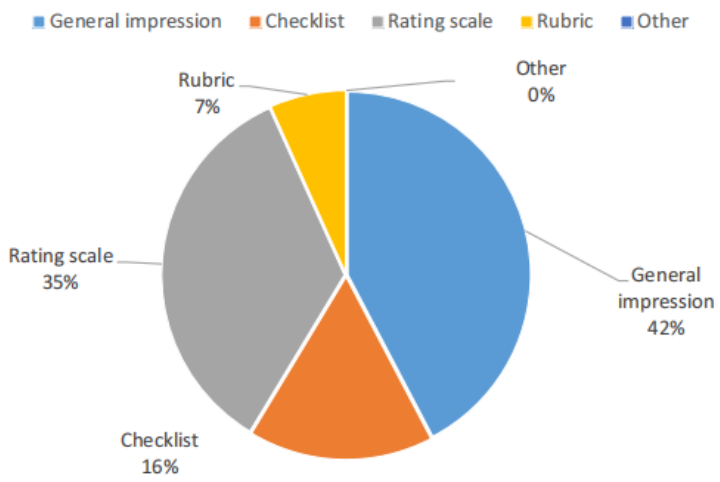

Figure 1. Assessment tools teachers applied

The quantitative data revealed that teachers were less likely to praise desired behaviors related to singing performance (M=3.30, SD=1.40). Instead, they frequently praised their own performance (e.g., ability or effort). The qualitative data confirmed the quantitative results that teachers frequently gave such personal praise. Only Ms. C claimed that praise should only be given when an expected performance was noted. In addition, the qualitative data showed that one teacher (Ms. F) gave students small gifts as external rewards. The quantitative data indicated that some teachers gave learners grades in their feedback in the form of a number or percentage, although this was rare (M=3.26, SD=1.094). The qualitative data confirmed the quantitative data which indicated that a few teachers graded their students with letters. Other teachers, such as Mr. I. and Mr. E., attached comments to their grades. The quantitative data revealed that teachers frequently showed students what to do (M=3.91, SD=1.005) and guided students to act upon the feedback in order to improve their learning (M=3.93, SD=0.961). Although the qualitative data did not directly address the frequency with which teachers guided students to respond to teacher feedback, it did reveal several strategies that teachers applied to guide students to respond to teacher feedback. Mr. D. mentioned two strategies: having individual conferences with the students to engage them in self-reflection and guiding the students to mark any existing problems or suggestions on their scores with different coloured pens immediately after receiving the teacher's feedback. Mr. E also shared his methods, that was using online software like WeChat to track students' practice. Nevertheless, more teachers mentioned the difficulty of systematically tracking students' response to feedback due to negative students' character, oversized classes and time constraints. For instance, Ms. F often invested some breaks or time after class to check if students have made progress, leading to increasing workload.

The quantitative data showed that while teachers had guided their students to identify their strengths and weaknesses in their performance, they rarely guided them to find strategies to improve their own work (M=3.37, SD=1.09). The qualitative data confirmed these results and showed that in many cases of self-assessment, teachers, rather than students, made suggestions for further improvement. One teacher (Ms. A) even assumed that students would recognize the errors in their performance through self-assessment and could naturally remedy such flaws by themselves without intervention.

5. Discussion

In this study, we have investigated how teachers conduct formative assessment in singing classes in Macao. The convergence of quantitative and qualitative data revealed that teachers in singing classes mainly use skill learning objectives, although knowledge objectives and dispositional objectives may also be used. The success criteria can be divided into two types: success criteria that relate to technical accuracy, and success criteria that relate to expressive qualities. Success criteria related to technical accuracy were more common than those related to expressive qualities. Cranmore and Wilhelm (2017) found similar results in their research. The explanation for this result was that most students in non-higher education in Macao were non-professional music students. Therefore, technical accuracy, which entails skills such as pitch and rhythm accuracy, might be challenging for these students. Consequently, teachers should strive for technical accuracy first, before adding expressive qualities. While eliciting learning evidence, teachers reported both questioning and performance assessment. As the main aim of singing teaching is to teach singing skills, performance assessment is applied as the primary assessment method that can take various forms. We also found that teachers varied their strategies for delivering feedback, including timing strategies and feedback complexity, depending on various instructional contexts. In

addition, we found that teachers applied far more teacher-directed formative assessment than student-directed formative assessment. This finding is consistent with previous studies, such as Wong (2014), Yan and Pastore (2022), and Holcomb (2019). Teacher-perceived factors that hindered student-directed assessment included the limitations of instructional time (Yan & Pastore, 2022), students' limited abilities (Yan & Pastore, 2022), and teachers' lack of confidence.

The strength of formative assessment is demonstrated in Strategies One, Three, Four, and Five. The convergence of quantitative and the qualitative data indicated that teachers applied multiple methods to clarify learning success criteria (Sang, 1987; Clarke, 2014; Hattie & Clarke, 2018). In terms of feedback specificity, teachers often pointed out the objects which had received the feedback, such as a person, a vocal part or the whole team (Martin, 2019). Also, teachers provided significantly more verbal feedback than written feedback. As verbal feedback was more timely than written feedback, it was more suitable to be used in singing lessons. Meanwhile, teachers also employed several effective methods while implementing self- and peer assessment. We examined only mixed convergence in terms of the amount of the two types of feedback: feed back and feed forward. At the process level, teachers gave more feed forward than feed back. At the task level, on the other hand, they gave more feedback than feed forward. Such difference can be explained by the different feedback levels. Therefore, these results may partially contradict Shaw's (2018) finding that "teachers gave too few 'next steps', not too many". Nevertheless, the self-reported data in this study might not accurately reflect the feedback teachers gave in real classes. Based on these findings, we recommend that teachers maintain these current strengths and build them on the higher level in future practice.

In contrast, weakness in implementing formative assessment was found in Strategies Two, Three, and Four. During the implementation of performance assessment, many teachers used general impressions rather than assessment tools to assess their students, which could have led to subjectivity and bias. Therefore, teachers should develop and utilize effective assessment tools such as checklists, rating scales, and rubrics when conducting performance assessments (Green & Hale, 2011; Valle, 2015; Pellegrino, 2015; Shaw, 2018). Moreover, assessment tools should be used 'formatively', that is, they should facilitate the generation of specific feedback, and support teachers in analysing data (Wesolowski, 2012). On this basis, effective adjustments to teaching can then be made. In addition, such tools should also be used for self- and peer-assessment (Valle, 2015; Hsia et al., 2016). The convergence of the quantitative the qualitative data confirmed that teachers provided multiple forms of evaluative feedback to their students, including grades, external rewards, and praise. Since it has been confirmed that evaluative feedback has little effect on student achievement, giving such feedback to students should be avoided (Shute, 2008; Brookhart, 2017; Wiliam, 2018). For instance, praise should only be given for the purposes of developing rapport (McPherson et al., 2022). Instead, teachers should provide students with specific and focused descriptive feedback (Brookhart, 2017). To achieve better specificity of feedback, teachers could: a) use many nouns and descriptive adjectives instead of pronouns; b) describe concepts or criteria related to learning tasks; and c) describe learning strategies that might be useful (Brookhart, 2017). As revealed by this study, teachers might avoid using vague words in their comments because such words might have multiple meanings. McPherson et al. (2022) also emphasized that teacher feedback should not be generalised, that is, teachers should specify how a goal can be achieved, rather than focusing on students' errors. Furthermore, the qualitative data enhanced the quantitative data by indicating that it might be difficult for teachers to systematically track students' responses. Therefore, teachers should be mindful that "giving does not mean receiving", that is, after giving students feedback, teachers should use effective strategies to track how they respond to the feedback (Wiliam, 2019). These strategies could improve current practices, such as checking students' understanding of feedback (McPherson et al., 2022), modelling the application of feedback (Leahy & Wiliam, 2014), and increasing dialogue instead of monologue (McPherson et al., 2022). In terms of self-assessment, teachers should not only guide students to critique their own work, but also to identify the 'next steps' and revise their own work (Andrea, 2010; Valle et al, 2016). On this basis, students' self-regulated skills could be developed so that they may become the owners of their own learning.

6. Implications

Although this study was conducted in a small territory (i.e. Macao), it examined the practice of formative assessment within music education in a systematic framework. The research results showed that teachers tended to use teacher-centered, rather than student-centered, assessment. These results could have been due to limited instructional time, pressure from public performances or competitions, or teachers' limited skills. Teachers' ability to implement formative assessment strategies also varied, and weaknesses were found in several of their strategies. This finding is consistent with the findings of previous studies that teachers could not implement every formative assessment strategy with equal effectiveness and have difficulty transferring their knowledge to classroom practice regardless of their teaching experience (Box et al., 2015; Johnson et al., 2019). On this basis, an obvious gap can be observed between the regulations or documents on formative assessment and its actual implementation, in which formative assessment functions as the main assessment method in Macao's education

system. Consequently, as several studies suggest, there is a need for continuous support and professional training for teachers (Yan & Cheng, 2015; Yan et al., 2021). Moreover, such professional training should be specific to the teaching areas of music. In other words, they should not only inform teachers about the concepts, characters, or general principles of formative assessment, but also contextualize it in singing teaching, such as developing and using assessment tools, providing descriptive feedback, increasing students' use of teacher feedback, and facilitating self- and peer-assessment. Additionally, future professional training should provide a welcoming environment for teachers and help them integrate the practice of formative assessment into their ongoing instruction, i.e., designing lesson plans and implementing them in the classroom (Wong, 2007). Accordingly, classroom observation and reflection should also be applied so that teachers can receive meaningful feedback for further improvement from various sources, such as colleagues, head teachers, school leaders, or outside experts.

Declaration of Conflicting Interests

The authors of this study declare that there are no potential conflicts of interest with respect to this article's research, authorship, and/or publication.

Funding

The authors received no financial support for this article's research, authorship, and/or publication.

References

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.

Black, P., & Wiliam, D., (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74.

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.

Box, C., Skoog, G., & Dabbs, J. M., (2015). A case study of teacher personal practice assessment theories and complexities of implementing formative assessment. *American Educational Research Journal*, *52*(5), 956-983.

Brookhart, S. M., (2017). How to give effective feedback to your students. Ascd.

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), 474-482.

Creswell, J. W., & Clark, V. L. P., (2007). *Designing and conducting mixed methods research*. Sage publications.

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.

Clarke, S., (2014). *Outstanding formative assessment: Culture and practice*. Hachette. UK.

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), 1-23.

Denis, J. M., (2018). Assessment in music: A practitioner introduction to assessing students. *Update: Applications of Research in Music Education*, *36*(3), 20-28.

Direcçãodos Serviços de Educaçãoe 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

Direcçãodos Serviços de Educaçãoe Juventude [DSEJ], (2017a). Junior Secondary School Music Curriculum Guide. Retrieved from http://www.dsej.gov.mo/crdc/guide/data/junior/Music_guide_c.pdf

Direcçãodos Serviços de Educaçãoe 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_2 8_2020e.pdf

Fitzpatrick, K. R., (2011). A mixed methods portrait of urban instrumental music teaching. *Journal of Research in Music Education*, *59*(3), 229-256.

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.

Green, S. K., & Hale, C. L., (2011). Fostering a lifelong love of music: Instruction and assessment practices that

make a difference. *Music Educators Journal*, *98*(1), 45-50.

Hattie, J., & Timperley, H., (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Hattie, J., & Clarke, S., (2018). *Visible learning: feedback*. Routledge.

Hodgen, J., & Marshall, B., (2005). Assessment for learning in English and mathematics: A comparison. *Curriculum Journal*, *16*(2), 153-176.

Holcomb, A., (2019). Assessment practices of American choral music educators. *The Oxford Handbook of Assessment Policy and Practice in Music Education*, Volume 2, 378-400.

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.

Johnson, C. C., Sondergeld, T. A., & Walton, J. B., (2019). A study of the implementation of formative assessment in three large urban districts. *American Educational Research Journal*, *56*(6), 2408-2438.

Leahy, S., & Wiliam, D., (2014). *Embedding formative assessment*. London, UK: Specialist.

Martin, L., (2020). Using synchronous formative feedback to facilitate student growth. *Music Educators Journal*, *107*(2), 51-57.

McPherson, G. E., Davidson, J. W., & Faulkner, R., (2012). *Music in our lives: Rethinking musical ability, development and identity*. Oxford University Press.

McPherson, G. E., Blackwell, J., & Hattie, J., (2022). Feedback in music performance teaching. *Frontiers in Psychology*, *13*, 1-11.

Pellegrino, K., Conway, C. M., & Russell, J. A., (2015). Assessment in performance-based secondary music classes. *Music Educators Journal*, *102*(1), 48-55.

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.

Scott, S. J., (2012). Rethinking the roles of assessment in music education. *Music Educators Journal*, *98*(3), 31-35.

Shaw, B. P., (2018). *Music assessment for better ensembles*. Oxford University Press.

Shute, V. J., (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189.

Tashakkori, A., & Teddlie, C., (2021). *Sage handbook of mixed methods in social & behavioral research*. SAGE publications.

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.

Wesolowski, B. C., (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, *98*(3), 36-42.

Wiliam, D., (2019). *Embedded formative assessment*. Solution tree press.

Wong, M. W., (2007). Assessment for learning and teacher development: The experience of three Hong Kong teachers. *Teacher Development*, *11*(3), 295-312.

Wong, M. W., (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.

Valle, C., (2015). Effects of criteria-referenced formative assessment on achievement in music. [Doctoral dissertation, State University of New York at Albany]. ProQuest Dissertations and Theses Global.

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.

Yan, Z., & Cheng, E. C. K., (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education*, *45*, 128-136.

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.

Yan, Z., & Pastore, S., (2022). Assessing teachers' strategies informative assessment: The teacher formative assessment practice scale. *Journal of Psychoeducational Assessment*, *40*(5), 592-604.

Appendix 1

Frequencies of Teachers' Implementation of Formative Assessment

Frequency of Implementing Formative Assessment: Scale 1 (not at all), 2 (once a semester), 3 (monthly), 4 (weekly), 5 (once or more times a day)

	N	Minimum	Maximum	Mean	Std. Deviation
I share the learning goal before students start working in singing class.	57	2	5	3.89	.880
The learning goal for the singing lesson is connected to local academic standards.	57	2	5	3.77	.945
I refer to the learning goal multiple times within the singing lesson.	57	1	5	3.46	.965
I share with students the criteria that will be used to determine their success in the singing lesson.	57	1	5	3.72	.978
I have students participate in developing the criteria for success in the singing lesson.	57	1	5	3.18	1.182
I provide demonstrations, models, or examples of singing when I articulate the criteria for success.	57	3	5	4.44	.732
I ask questions during the singing lesson to assess the whole group's understanding of singing.	57	2	5	4.07	.863
I ask questions during the singing lesson to assess individual students' understanding of singing.	57	2	5	4.00	.926
I adjust the instructions within the singing lesson based upon student responses.	57	1	5	3.93	.997
I ensure that the pace of the singing lesson provides adequate down time for students to respond to questions.	57	2	5	4.00	.802
I use follow-up questions when engaging students in discourse.	57	2	5	4.26	.791
I assess the learners' singing by utilizing both live and in-class singing exams.	57	2	5	3.82	.966
I assess the learners' singing by both observing and judging their in-class singing performance.	57	2	5	4.14	.833
I assess the technical accuracy (e.g., pitch accuracy, rhythm accuracy, pronunciation, and vocal technique) when I assess learners' singing.	57	2	5	4.32	.985
I assess the musical expression (e.g., dynamics, timbre, phrasing, articulation, composer/stylistic intent, etc.) when I assess learners' singing.	57	2	5	4.09	.931
I give learners a grade in my feedback in the form of a number or a percentage.	57	1	5	3.26	1.094
I tell students what they have *not* achieved with specific references to their learning.	57	1	5	3.72	1.114
I tell students what they *have* achieved with specific references to their learning.	57	2	5	3.74	.936
I specify a better or different singing strategy that they can use for improvement.	57	1	5	4.14	1.008
I point out the objects that need improvement such as a person, a vocal part, or a whole team.	57	2	5	4.25	.851
I provide praise related to singing performance, not to the learners themselves (i.e., their ability or effort) at the self level	57	1	5	3.30	1.401

I provide immediate, direct feedback during the process of a repertoire.	57	1	5	3.88	1.135
I give delayed feedback that is shared several minutes following the completion of a repertoire.	57	1	5	3.98	1.061
I provide corrective information (such as verifying “right” or “wrong”, providing correct responses, error flagging, or “try again”) in the singing class through explanation or demonstration.	57	1	5	4.18	.869
I offer commentary (such as hints, cues, or prompts) that guides students to independent observations and choices.	57	1	5	4.19	.811
I provide verbal feedback to the learners.	57	1	5	4.39	.840
I provide written feedback to the learners.	57	1	5	2.68	1.242
I show students what they need to do in order to improve their learning based on assessment results.	57	1	5	3.91	1.005
I guide students to act on assessment feedback information to improve their learning.	57	1	5	3.93	.961
I use student self-assessment in singing lessons.	57	1	5	3.25	1.138
I guide students to identify strengths and weaknesses in their own singing performance.	57	1	5	3.79	.940
I guide the learners to indicate both the locations and the dimensions of their weaknesses in their singing performance.	57	1	5	3.70	.906
I ask the learners to record themselves and then listen to these recordings to identify, articulate, and correct any mistakes.	57	1	5	3.42	1.253
I guide students to identify strategies that will improve their own work.	57	1	5	3.37	1.096
I use evidence generated through student self-assessments to inform future teaching and learning.	57	1	5	3.42	1.133
I help students develop their self-assessment skills.	57	1	5	3.49	.984
I use student peer-assessment in singing lessons.	57	1	5	3.35	1.142
I guide students to provide feedback to help their peers improve.	57	1	5	3.54	1.103
I use evidence generated through student peer-assessments to inform future teaching and learning.	57	1	5	3.25	1.184
I provide the learners with sentence starters (e.g., “I like the way you...”, “You did an excellent job of ...”, “I was surprised that...”, and “I do not understand”) to prompt them.	57	1	5	3.68	1.198
I monitor the peer assessment process by circulating among the pairs, giving feedback, coaching, and sequencing activities, as necessary.	57	1	5	3.58	.999
I teach students to engage in peer feedback processes.	57	1	5	3.46	.946
Valid N (listwise)	57				