ARTICLE

Proposing and piloting a criterion- and standard-based assessment framework in teaching Cantonese operatic singing in Guangdong, China

Yue Luo[1] and Bo-Wah Leung[2]*

[1]Research Centre for Transmission of Cantonese Opera, The Education University of Hong Kong and [2]Department of Cultural and Creative Arts, The Education University of Hong Kong
*Corresponding author: Bo-Wah Leung, Email: bwleung@eduhk.hk

Abstract

Teaching traditional art forms in schools and the community has proven an effective way of ensuring the transmission of traditional culture. But due to the lack of valid and normative assessment guidance, the assessments of Cantonese operatic singing are still developing, impairing the instructiveness in its teaching and learning and restraining the education and development of this traditional genre within contemporary society. Accordingly, to guide the design of a holistic assessment, the revised Bloom's taxonomy was adopted to form a theoretical framework from which a criterion- and standard-based assessment framework with four domains has been proposed. Four teachers and 24 students of Cantonese opera institutions in Guangdong, China, were invited to pilot the assessment framework. Afterwards, semi-structured interviews were conducted from which two research questions were answered satisfactorily: 1) proposing a criteria- and standard-based assessment framework is necessary, for the traditional assessment practice is weak in guiding teaching and learning and 2) positive feedback supported the proposed assessment framework facilitates the teaching and learning of Cantonese operatic singing.

Keywords: Cantonese opera; formative assessment; criteria-based assessment; standard-based assessment; music education in China

Introduction

Known as *Yueju* (粵劇), Cantonese opera is one of the Chinese music traditions included on the UNESCO Representative List of Intangible Cultural Heritage of Humanities (2009). The history of this art form can be traced back to the Southern Song Dynasty (1127~1279) (Chan, 1999) or Ming Dynasty (1638~1644) (Leung, 1982). It is still popular worldwide, especially in its heartlands of south China: Guangdong and Guangxi Provinces, Hong Kong and Macau. Like some other folk arts in East Asia, Cantonese opera was transmitted using apprenticeship before the mid-twentieth century, which featured a quasi-parental relationship, an oral approach and informal learning (Leung, 2015). The apprenticeship system in China is considered to be dated back to the Qin dynasty (200 BC) (Zhang & Cerdin, 2020), which is also recognised as the first trace of vocational training in China (Zai et al., 2018) and became the origin mode of transmission and education, widely used in handicrafts, art, *Kungfu* and Chinese medicine education. Although conservatory tradition and community training had taken the place of apprenticeship and become the main transmission and education form of Cantonese opera in the mid-twentieth century (Leung, 2015), the conventional practice of apprenticeship still influences its assessment in daily teaching and learning. Due to the lack of valid and normative assessment guidance, the traditional assessments of Cantonese operatic singing are thought to be opaque, incomplete, subjective and
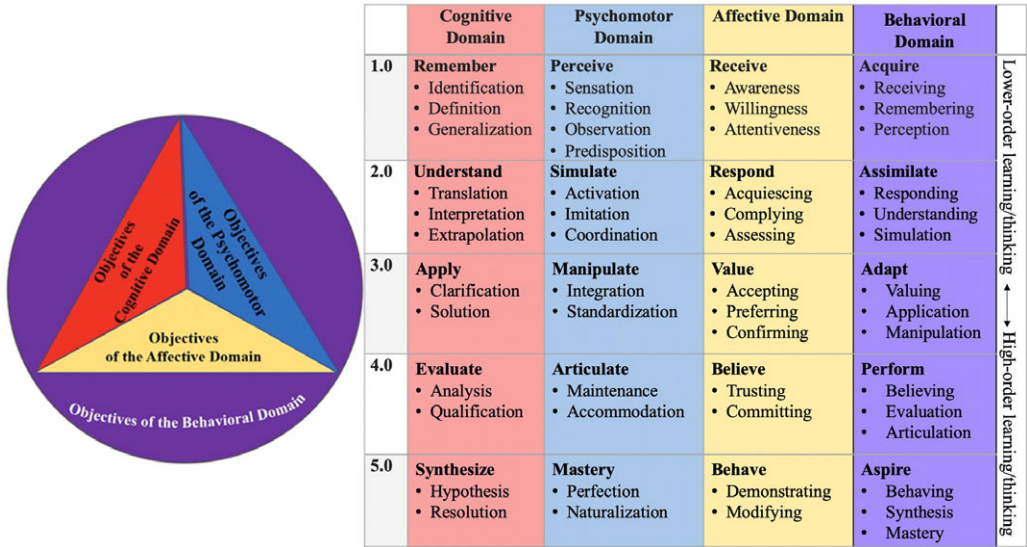
Figure 1. Theoretical framework for assessments based on revised Bloom's taxonomy (Luo & Leung, 2022).

arbitrary, impairing the instructiveness in its teaching and learning and restraining the education and development of this traditional genre within contemporary society (Luo & Leung, 2022). Thus, based on a holistic theoretical framework stemming from the revised Bloom's taxonomy, a criterion- and standard-based assessment framework was proposed and piloted in two institutions: one is the Cantonese Opera School, the cradle of future professional artists of Cantonese opera, representing the conservatory tradition. The other is a Children's Palace in Guangzhou, China, which is a public facility in China where children engage in extra-curricular activities, representing community training.

Background

The proposed theoretical framework

Bloom's Taxonomy is a set of cumulative hierarchical models that can be used to categorise educational learning objectives into levels of specificity and complexity in different learning domains (Bloom, 1956). Hierarchies are assigned to each of these domains that correspond to different levels of learning or thinking, from low to high, making discussion of educational objectives easier, guiding instruction and allowing useful assessment (Krathwohl, 2002). It facilitates careful inspection and ultimately reinforces the connections between the components of education: curriculum, instruction and assessment (Airasian & Miranda, 2002). Thus, it is considered a useful tool for stating objectives, building curricula and constructing and testing evaluation procedures (Anderson & Sosniak, 1994). Making good use of this tool contributes to teasing out suitable assessment criteria and standards so that reliable assessments can be ensured.

To improve the education and transmission of Chinese *Xiqu*, a theoretical framework derived from the revised Bloom's Taxonomy is proposed to ensure assessment quality (Luo & Leung, 2022). The four domains of this theoretical framework with their five respective hierarchies from low to high can be synthesised in Figure 1.

To better demonstrate the five hierarchical structures of every domain shown in Figure 1, three-dimensional models were diagrammatized from different angles shown in Figure 2.

Based on this theoretical framework, the current study strives to propose a criterion- and standard-based assessment framework and conduct a pilot implementation on a small scale in the context of Cantonese operatic singing in China to seek participant feedback for further improvement before a larger-scale study.
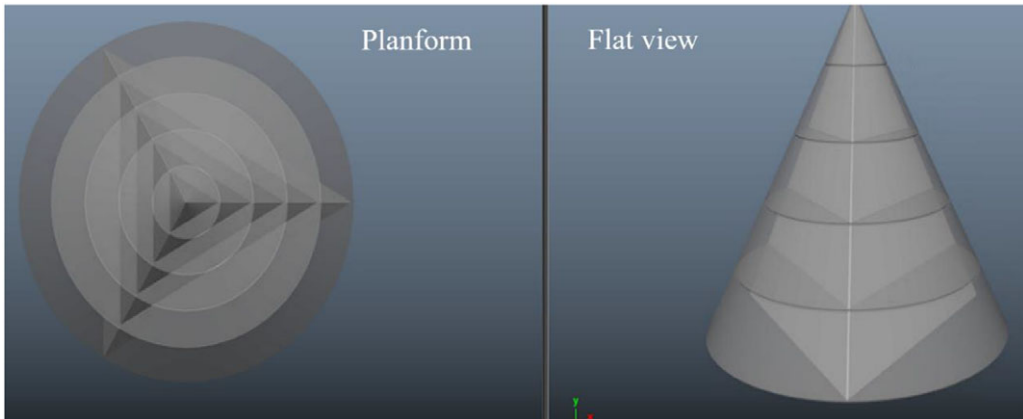
Figure 2. Three-dimensional model of the theoretical framework.

Proposing a criterion- and standard-based assessment framework

The criterion- and standard-based assessment framework is an outcome of a doctoral study based on a holistic theoretical framework stemming from the revised Bloom's taxonomy comprised of the Cognitive Domain, Psychomotor Domain, Affective Domain and Behavioural Domain (Luo & Leung, 2022). Teachers can judge students' levels and awareness of relevant domains from student answers or performance, corresponding to the range hierarchically from low to high in the theoretical framework. Thus, these hierarchies become definitions of different levels when setting the assessment criteria. This permits the criteria and weightings to be fixed with the users, including teachers or assessors. But in the current study, weightings were set as equal divisions to focus on the assessment system's overall structure.

Intellectual abilities and skills (Cognitive Domain)

This domain involves intellectual abilities and skills (Bloom, 1956), manifesting themselves in open-ended questions and answers about relevant knowledge (Anderson & Krathwohl, 2001). In the context of Cantonese operatic singing, this domain is suited to assessing intellectual subjects or content, such as the history, the general knowledge, or the structure of Cantonese opera. To test the Cognitive Domain of the proposed assessment system in the pilot implementation, three questions were set assessing relevant general knowledge of Cantonese opera (see Table 1).

Users were requested to employ the assessment framework when assessing students based on an understanding of the proposed hierarchies. Each level from 1.0 to 5.0 was divided into four levels represented by grids for further subdivisions. The further to the right, the higher the level represented. For each line from 1.0 to 5.0, users were required to select only one of the grids according to the student's performance. Algorithms were adopted in the assessment framework to avoid human factors and ensure objectivity. Since every line was assigned a hundred points and there were 20 grids to each line, a grid in the same line was assigned five scores. This practice was also adopted in the Psychomotor Domain and Behavioural Domain.

Physical abilities and skills (Psychomotor Domain)

This domain involves the manipulative motor-skill area (Bloom, 1956), manifesting itself in physical abilities and skills (Hauenstein, 1998). Although the following assessment framework might be used to assess any physical ability, skill in Cantonese operatic singing was chosen as the subject of the pilot implementation and used to verify the assessment framework of the psychomotor

Table 1. Assessment Framework of the Structure of Cantonese opera (Cognitive Domain)

Standards / Items	Low — 1.0 Remember				2.0 Understand				3.0 Apply				4.0 Analyze				5.0 Synthesize — High				Weightings
Q1	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	1/3
Q2	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	1/3
Q3	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	1/3
																			Score:		0.00

Table 2. Assessment Framework of Cantonese Operatic Singing Skill (Psychomotor Domain)

Name:									Grade:											
Standards / Criteira	Low — 1.0 Perceive				2.0 Simulate				3.0 Manipulate				4.0 Articulate				5.0 Master — High				Weightings
Intonation	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	20%
Rhythm	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	20%
Breath	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	20%
Articulation	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	20%
Timber	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	20%
																			Score:		0

domain. According to a consensus drawn from the teacher participants, an assessment framework for Cantonese operatic singing skills was established, as shown in Table 2.

Feelings, emotions and attitudes (Affective Domain)

This domain involves human beings' feelings, emotions and attitudes, which includes how one deals with things emotionally, such as feelings, appreciation, values, motivations, enthusiasms and attitudes (Krathwohl et al., 1964; Gronlund & Brookhart, 2009). Psychologists have identified numerous constructs that reflect affective characteristics. The current study adopted the ones McCoach and his fellows advocated, including attitudes, self-efficacy, values, motivation and interest. McCoach et al. (2013) stated that "in general terms, attitudes were described as feelings towards some object; self-efficacy was referred to as a self-appraisal of capability; values reflected enduring beliefs; motivation can denote both external and internal states that drive us in a particular direction, and interests reflected preferences for particular activities" (p. 25). These affective characteristics are explicitly specified and applicable to Cantonese opera's educational context. Thus, the questions for assessing the affective domain in the current study were developed around these five affective constructs with the teachers.

There are three main dimensions of the affective domain: 1) recorded data, 2) self-reported data and 3) observational data (Geisert, 1972). Since using multiple methods to collect information on various characteristics from multiple sources is deemed particularly important in affective assessment (Oakland, 1997), each of the above three main dimensions was incorporated.

Recorded data, the first dimension of the affective domain, are widely collected in schools, such as absenteeism, tardiness, homework performance, discipline, etc. Because these data have been

Table 3. Survey of Student Perception on Learning (For professional students)

Standards / Criteria	Strongly agree ⟶ Strongly disagree				
	1	2	3	4	5
1. I think it is important for my career to learn this course well.					
2. I have a feeling of powerlessness when studying this course					
3. I am very interested in learning this course.					
4. I am willing to spend more time and energy on this course.					
5. I am confident that I can master the knowledge and skills of this course.					
6. This course cannot hold my attention at all.					
7. To learn this course well, I have taken the initiative to consult teachers or predecessors.					
8. I think this course is useful for the transmission of traditional Chinese culture, which gives me a sense of mission and pride.					
9. I am not satisfied with my performance in this course.					
10. I enjoy the class.					

routinely gathered and documented in the Cantonese Opera School and the Children's Palace, it was decided to follow existing practices in these institutions rather than establish new practices.
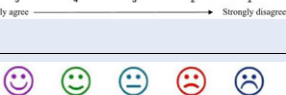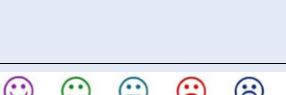
Self-reporting, the second dimension of the affective domain, is considered the most common and direct measure of affective traits (McCoach et al., 2013), making it the primary approach for collecting affective assessment data. Multiple instruments can be adopted when conducting an affective assessment, such as the Likert scale, semantic differential scale, Q-SORT and the like (Geisert, 1972; Hall, 2011). The most widespread practice in the affective instrument is to "ask respondents to select one response from an ordinal series of response options" (McCoach et al., 2013, p. 40) so that a consistent "frame of reference for all the respondents" can be provided (Weisberg et al., 1996, p. 84). Therefore, the current study employed a 5-point Likert scale as the instrument for student self-reporting. As respondents prefer close-ended questions to open-ended questions (Dillman et al., 2009), close-ended questions were used to collect both self-reported and subsequent observational data.

The questions referred to some examples in the post-experimental Intrinsic Motivation Inventory (Center for Self-Determination Theory, n.d.). To ensure reliability, cross-checking was considered with an adequate number of questions (Reid, 2006). Consequently, two related questions aimed at one affective construct were set to be in positive and negative ways of saying. The congruent relationships between affective constructs and the sequence numbers of questions were 1) value: 1 & 8; 2) attitude: 2, 10; 3) interest: 3, 6; 4) motivation: 4,7; 5) self-efficacy: 5, 9. Given the various positions of professional students (professional cultivation) and amateurs (hobby), questions varied slightly in accordance. The question statements in Table 3 received approval from teachers who participated in the pilot implementation.

Since student ages in Children's Palace range from 7 to 14, images were introduced into the options to make it easier for younger respondents to employ (see Table 4).

Despite the multiple advantages of self-reporting, an overarching disadvantage is its credibility (Paulhus & Vazire, 2007). Participants might distort their responses (Paulhus, 1991), and some might even be concerned about how the instrument designers coded or measured deviations unknowingly generated by respondents (Wilson et al., 2000; Greenwald & Banaji, 2003;

Table 4. Survey of Student Perception on Learning (For amateur students)

	Name:	Grade:
1	I think this is an important course.	
2	I have a feeling of powerlessness when studying this course.	
3	I would describe this course as very interesting.	
4	I put a lot of effort into this.	
5	I am confident that I can master the knowledge and skills of this course.	
6	I thought this was a boring course.	
7	It was my family who wanted me to learn Cantonese opera rather than my own decision.	
8	I believe learning this course could be beneficial to me.	
9	I am satisfied with my performance in this course.	
10	I enjoyed the class.	

Each item uses a 5-point scale with face icons: 5 Strongly agree — 4 — 3 — 2 — 1 Strongly disagree.

Table 5. Observational Assessment of Student Attitude and Psychology in Learning (For professional students)

Standards / Criteria	Lower-level ←——————————————→ Higher-level				
	1.0 Receive	2.0 Respond	3.0 Value	4.0 Believe	5.0 Behave
1. Regarding the importance of this subject, the student's behaviour is:					
2. When confronted with learning difficulties thereby receiving encouragement and instruction of solutions from the teacher, the student's behaviour is:					
3. When the teacher stimulates the student's interest in learning in certain approaches, the student's behaviour is:					
4. When encouraged to devote more time and energy to the course, the student's behaviour is:					
5. Regarding the confidence to master the knowledge and skills of this course, the student's behaviour is:					
6. In the class, the student manifests himself as:					
7. Regarding the extra effort that may be required to learn the course well, the student's behaviour is:					
8. Regarding the relevance of Cantonese opera to the inheritance of traditional Chinese culture, the student's behaviour is:					
9. Regarding the performance in the class, the student's behaviour is:					
10. Regarding the emotions in the class, the student's behaviour is:					

Blanton & Jaccard, 2006). Therefore, multiple sources for the assessment of the affective domain were necessary.

Besides the two dimensions above, the third dimension is observational data collected from a target population and evaluated by teachers based on the hierarchies of the affective domain. This included 1.0 Receive, 2.0 Respond, and the like. To collect the same information for every question from every student and teacher, the questions corresponded to those in the student self-reported questionnaire. Thus, the congruent relationships between affective constructs and the sequence numbers of the questions were 1) Value: 1 & 8; 2) Attitude: 2, 10; 3) Interest: 3, 6; 4) Motivation: 4, 7 and 5) Self-efficacy: 5, 9 (see Table 5).

Comprehensive Performance (Behavioral Domain)

This domain embraces an integration of cognitive, affective and psychomotor domains (Hauenstein, 1998). An assessment framework for the singing class was established based on a consensus agreed with the teacher participants, as shown in Table 6.

A pilot implementation was conducted to examine the feasibility and identify potential problems and deficiencies in the proposed assessment framework. Pilot implementation is defined as "a field test of a properly engineered, yet unfinished system in its intended environment, using real data and aiming – through real-use experience – to explore the value of the system, improve or assess its design, and reduce implementation risk" (Hertzum et al., 2012, p. 314).

Table 6. The Assessment Framework of the Singing Class in Cantonese Operatic Singing (Behavioral Domain)

Name:					Grade:	
Standards \\ Criteria	Low ──────────────────────────► High					Weightings
	1.0 Acquire	2.0 Assimilate	3.0 Adapt	4.0 Perform	5.0 Perfect	
Technical requirements						
Intonation	○○○○	○○○○	○○○○	○○○○	○○○○	1/6
Pulse & Rhythm	○○○○	○○○○	○○○○	○○○○	○○○○	1/6
Cantonese pronunciation	○○○○	○○○○	○○○○	○○○○	○○○○	1/6
Changqiang	○○○○	○○○○	○○○○	○○○○	○○○○	1/6
Artistic requirements						
Artistic expression in line with the repertoire	○○○○	○○○○	○○○○	○○○○	○○○○	1/6
The appearance in costume and makeup	○○○○	○○○○	○○○○	○○○○	○○○○	1/6
					Score:	0

Research question

The current study focuses on the following research questions: 1) Why propose criteria- and standard-based assessment framework for Cantonese operatic singing? 2) How well does the proposed assessment framework facilitate the teaching and learning of Cantonese operatic singing?

Methodology

Surrounding the research questions, the current study consisted of two phases. In Phase 1, a series of pilot implementations was conducted to: 1) evaluate the usefulness and usability of the proposed assessment framework and 2) identify necessary or desirable changes in work organisation and processes. Detailed elaboration and explanations of the proposed assessment framework and the procedure of the pilot implementation were offered to the participants before the pilot was implemented. Consensus-based comprehension was also ensured before the assessment was launched. Teacher participants were asked to assess the student participants' general knowledge of Cantonese opera, their Cantonese operatic singing skills, and their singing class performance, which corresponded to, respectively, the cognitive domain, psychomotor domain and behavioural domain. Both online student self-reports and teacher observational assessments were conducted to assess the affective domain. All assessment results were tabulated and distributed to the participants to be reviewed before the semi-structured interviews. This was to 1) allow comparison with the results generated from the traditional assessment and 2) provide information and lay a foundation for communication during the interviews.

In Phase 2, semi-structured interviews were conducted with the participants to collect feedback and suggestions. Since the amateur students from Children's Palace were aged from only 7 to 14, their parents were also invited to accompany their children. This had two purposes: 1) to meet ethical considerations and 2) to collect the parents' opinions. Interview questions included the following:

1) What is the current situation surrounding the assessment of Cantonese operatic singing?
2) For teacher-participants: What are your thoughts after applying this assessment framework?
   For student participants: What do you think about receiving the assessment results of this assessment framework?

3) Are there any distinct gaps between the assessment results and the average academic performance?
4) What are the merits and demerits of this assessment framework?
5) How does the proposed assessment framework facilitate teaching and learning?
6) What functional or practical problems in the assessment framework or work organisation and processes need to be solved before application?

Data collection

Purposive sampling was used to locate eligible participants who were the most representative and ensured that information-rich interviewees were selected (Cohen et al., 2018). Several veteran teachers were invited to be participants to examine the feasibility of the proposed assessment framework. Elderly teachers proved reluctant to join the pilot implementation. One middle-aged teacher from Children's Palace (code name: T1), representing community training, and three younger teachers from the Cantonese Opera School (code name: T2, T3 and T4), representing the conservatory tradition, showed strong interest in becoming teacher participants in the study.

Stratified sampling (Arnab, 2017) was adopted to select six categories of student participants, including top students, average students and underachievers in senior and junior grades. Teachers selected these six categories of students based on their daily academic performance. Thus, 24 student participants were assessed using the proposed assessment framework and joined the online Survey of Student Perception on Learning to engage in self-reporting of the affective domain.

The teacher participants were encoded as T1, T2, T3, and T4. Accordingly, the student participants were encoded S1 to S6 and attached to their teacher code names (see Table 7). For example, the first student of T1 was encoded T1-S1.

Data analysis

Thematic analysis is a systematic method employed to identify, organise and extract insightful themes across the data sets obtained from the semi-structured interviews (Clarke & Braun, 2014). The current study adopted a six-step approach to thematic analysis proposed by Braun and Clarke (2006): 1) Familiarisation with the data; 2) Generating initial codes; 3) Searching for themes; 4) Reviewing potential themes; 5) Defining and naming themes and 6) Producing the report. The first author conducted the thematic analysis, and the second author reviewed it as triangulation. To promote the effectiveness and efficiency of coding and analysing and to create data visualisations, MAXQDA, a world-leading software for qualitative data analysis, was employed.

Using MAXQDA, 587 codes and eight nodes were extracted from the text coding of the semi-structured interviews. Two themes were deduced, including "Traditional assessments are weak in guiding teaching & learning" and "Proposed assessment framework facilitates teaching & learning" (Tables 8 and 9).

Results

This section presents the findings from the assessment process and the results of the analysis of the qualitative interview data, from which rich insight can be derived.

Assessment results

Selected by their teachers using stratified sampling, the student participants were subsumed into six categories according to regular academic performance, including top students, average students and underachievers in senior grades and junior grades. The teachers made judgments of students' performance or answers by selecting the grids corresponding to the standards of every

Table 7. Codes for teacher and student participants in Phase II

Location	Teacher	Student
Children's Palace	T-1	T1-S1
		T1-S2
		T1-S3
		T1-S4
		T1-S5
		T1-S6
Cantonese Opera School	T-2	T2-S1
		T2-S2
		T2-S3
		T2-S4
		T2-S5
		T2-S6
	T-3	T3-S1
		T3-S2
		T3-S3
		T3-S4
		T3-S5
		T3-S6
	T-4	T4-S1
		T4-S2
		T4-S3
		T4-S4
		T4-S5
		T4-S6

criterion in the assessment framework. It not only formed progress bars, representing the levels of the students' performance towards every criterion but also triggered the algorithms set in Microsoft Excel that eventually generated a total point. As an example, the assessment results of one of the students are given in Figures 3–6.

The teachers distributed the assessment results for the four domains to every student to be reviewed. The answers to the Online Survey of Student Perception on Learning were exported from the database backend, where the survey data were stored and retrieved. It would be forwarded to the teachers for review (see Figures 7 and 8).

Unusual answers were highlighted for further interpretation. For example, one of the students (code name: T2-S1) chose 5 on the Likert Scale, denoting "strongly agree" as their answer to Q6: "This course cannot hold my attention at all" and chose 2, denoting "disagree" as their answer to Q10: "I enjoy the class". This seemingly reflected a disturbing learning status. The issue was later raised in the semi-structured interviews with the student and their teacher.

The total points allocated in the psychomotor domain, cognitive domain and performance domain of each student were tabulated for teacher participants and the researcher's review.

Table 8. Theme 1: Traditional assessment practice is weak in guiding teaching and learning

Nodes	Codes	Number of Codes
1. Characteristics		
	Assess without criteria	17
	Test score manipulation	37
	Avoid failure rate	19
	Protect students' self-esteem	18
	Affected by human factors	13
	Oral evaluation	33
	Assess by intuition or impression	17
2. Advantages		
	Habitually practice	2
	Pass rate can be controlled and ensured	3
3. Disadvantages		
	Lack of instructiveness	8
	General and Implicit	24
	Incomplete and non-normative	13
	Lack of objectivity and fairness	25
	Oral evaluation is apt to be forgotten	14
	Untransparent	3
4. Perceptions		
	Poor quality in assessment impairs education and transmission	17
	Advocate for impartial assessment	6
	Unconvincing assessment does not help guide learning	18
	Failing in assessment might hurt the students' initiative or prospect	4

The assessment results from the pilot implementation generally match the usual academic performances (See Table 10).

From Table 10, it is clear that the assessment results are almost in line with the regular academic performance of students, except for 1) student T3-S1, whose Q&A assessment result was much higher than average and top students at the same grade and 2) student T3-S2 whose Cantonese operatic singing skills assessment result was a little higher than the top student at the same grade. This unevenness became one of the topics discussed in the subsequent semi-structured interviews with Teacher No.3.

Whether the above assessment results were acceptable to the students and teachers and whether there was a distinct gap between the assessment results and normal academic performances became a focus of the subsequent semi-structured interviews.

Thematic networks and their connections from the semi-structured interviews

The thematic networks and the relations between the codes and nodes under these themes were extracted from the semi-structured interviews and illustrated in Figure 9. Three types of relations were proposed amongst the thematic networks, including "associated with" (denoted by the

Table 9. Theme 2: Proposed assessment framework facilitates teaching and learning

Nodes	Codes	Number of Codes
1. Characteristics		
	Criteria- and standard-based	32
	Comprehensive and systematic	19
	In written form	11
2. Disadvantages		
	Teachers' workload might be increased	4
	Time is needed to adapt to the new assessment system	2
	Not found yet	14
3. Advantages		
	Instructive and efficient	14
	Easy-start and user-friendly	16
	More detailed feedback	14
	Normative and all-sided	7
	Clear and transparent	26
	Explicit assessment objectives in advance	20
	Documentable	12
	Objective and convincing	21
4. Perceptions		
	Assessment results after pilot implementation	
	Coincide	15
	Lower	13
	Preference for assessment systems	
	Contradiction	3
	Approve of the proposed assessment framework	22
	Better than the traditional assessment	19
	Good supplement to the three-level promotion mechanism	2
	Elder teachers might oppose the new assessment framework	1
	Advocate for wider utilization	9

yellow arrows), "influenced by" (denoted by the red arrows) and "solved by" (denoted by the blue arrows).

In Theme 1, "traditional assessments are weak in guiding teaching & learning", we noticed four characteristics of traditional assessment practice, including "assess without criteria", "oral assessment", "assess by intuition or impression" and "test score manipulation" (See Figure 10).

Three pairs of interactions were detected in the first kind of relationship ("associated with") in Theme 1:

1) As we can see in the above schematic, "test score manipulation" is the most salient problem in the traditional assessment practice of professional education of Cantonese opera. It was

Code Name: T2-S6   **Grade: Senior**

Standards / Items	Low — 1.0 Perceive	2.0 Simulate	3.0 Manipulate	4.0 Articulate	5.0 Master — High	Weightings
Intonation				● (4.0)		20%
Rhythm					●	20%
Breath					●	20%
Articulation					●	20%
Timbre					●	20%

Score: 88

Figure 3. An example of the assessment result of Cantonese operatic singing skill.

Code Name: T2-S6   **Grade: Senior**

Standards / Items	Low — 1.0 Remember	2.0 Understand	3.0 Apply	4.0 Analyze	5.0 Synthesize	Weightings
Q1					●	1/3
Q2					●	1/3
Q3					●	1/3

Score: 86.67

Figure 4. An example of the assessment results of Q&A.

Code Name: T2-S6   **Grade: Senior**

Standards / Items	Low — 1.0 Acquire	2.0 Assimilate	3.0 Adapt	4.0 Perform	5.0 Perfect	Weightings
Technical requirements						
Intonation					●	1/6
Pulse & Rhythm					●	1/6
Cantonese pronunciation					●	1/6
Changqiang					●	1/6
Artistic requirements						
Artistic expression in line with the repertoire					●	1/6
The appearance in costume and makeup					●	1/6

Score: 84.17

Figure 5. An example of the assessment results of singing class in Cantonese operatic singing.

Code Name: T2-S6   Grade: Senior

Criteria \ Standards	Lower-level → Higher-level				
	1.0 Receive	2.0 Respond	3.0 Value	4.0 Believe	5.0 Behave
1. Regarding the importance of this subject, the student's behavior is:	○	○	●	○	○
2. When confronted with learning difficulties thereby receiving encouragement and instruction of solutions from the teacher, the student's behavior is:	○	○	○	●	○
3. When the teacher stimulates the student's interest in learning in certain approaches, the student's behavior is:	○	○	●	○	○
4. When encouraged to devote more time and energy to the course, the student's behavior is:	○	○	○	●	○
5. Regarding the confidence to master the knowledge and skills of this course, the student's behavior is:	○	○	○	●	○
6. In the class, the student manifests himself as:	○	○	○	●	○
7. Regarding the extra effort that may be required to learn the course well, the student's behavior is:	○	○	○	○	●
8. Regarding the relevance of Cantonese opera to the inheritance of traditional Chinese culture, the student's behavior is:	○	○	○	○	●
9. Regarding the performance in the class, the student's behavior is:	○	○	○	●	○
10. Regarding the emotions in the class, the student's behavior is:	○	○	○	○	●

Figure 6. An example of the observational assessment towards student attitude and psychology in learning.

mentioned 37 times in the interviews, reflecting a widespread phenomenon among the interviewees and a noteworthy characteristic.

2) "In traditional assessment, we usually make adjustments so that all students can pass the examinations. Otherwise, it would be vexatious for both students and teachers: If students get an F, they must take part in make-up examinations until they eventually pass. Otherwise, they cannot graduate. So, unlike the scoring in competitions, even though the students do not do well in the examination, as long as they are not far too poor, we would still give them at least a passing score to ensure the pass rate. For example, Table 11 is the transcript of one of the classes I taught at my school. Even the worst student got a passing score. And to distinguish different levels of students' academic performance, the scores of the other better students would be raised even higher correspondingly." (T2)

Teacher	Student	1. I think this is an important course.	2. I have a feeling of powerlessness when studying this course.	3. I would describe this course as very interesting.	4. I put a lot of effort into this.	5. I am confident that I can master the knowledge and skills of this course.	6. I thought this was a boring course.	7. It was my family who wanted me to learn Cantonese opera rather than my own decision.	8. I believe learning this course could be beneficial to me.	9. I am not satisfied with my performance at this course.	10. I enjoyed the class.
	T1-S1	4	1		4	5		2	5	2	5
	T1-S2	5	5		5	5		5	4	2	5
T1	T1-S3	4	2		5	5		1	4	2	5
	T1-S4	5	2		4	5		5	5	2	5
	T1-S5	5	2		4	5		2	5	1	5
	T1-S6	5	1		5	5		2	4	1	5

Likert scale:	1	2	3	4	5
	Strongly disagree	→	→	→	Strongly agree

Figure 7. Answers to online survey of student perception on learning (for amateur students).

Teacher	Student	1. I think it is important for my career to learn this course well.	2. I have a feeling of powerlessness when studying this course	3. I am very interested in learning this course.	4. I am willing to spend more time and energy on this course.	5. I am confident that I can master the knowledge and skills of this course.	6. This course cannot hold my attention at all.	7. To learn this course well, I have taken the initiative to consult teachers or predecessors.	8. I think this course is useful for the transmission of traditional Chinese culture, which gives me a sense of mission and pride.	9. I am not satisfied with my performance in this course.	10. I enjoy the class.
	T2-S1	5	3	5	5	4	5	4	5	3	2
	T2-S2	4	1	5	5	4	1	5	4	2	5
T2	T2-S3	5	3	4	4	4	1	4	5	1	5
	T2-S4	4	4	5	4	4	3	4	5	4	3
	T2-S5	4	3	5	5	4	3	4	5	1	4
	T2-S6	5	4	5	5	5	1	4	5	2	2
	T3-S1	5	1	5	5	5	1	5	4	1	5
	T3-S2	5	1	5	5	5	2	4	5	1	5
T3	T3-S3	5	1	5	5	5	1	5	5	1	5
	T3-S4	5	1	5	5	5	1	5	5	2	4
	T3-S5	5	4	5	5	4	5	3	4	5	2
	T3-S6	5	1	5	5	5	1	5	5	1	4
	T4-S1	5	1	5	5	5	1	4	4	1	5
	T4-S2	5	1	5	5	4	1	4	5	2	5
T4	T4-S3	5	1	5	4	5	1	4	5	2	5
	T4-S4	5	1	5	5	5	1	5	4	1	5
	T4-S5	4	3	5	4	4	2	5	5	1	4
	T4-S6	5	5	5	5	5	3	3	3	4	3

Likert scale:	1	2	3	4	5
	Strongly disagree	→	→	→	Strongly agree

Figure 8. Answers to online survey of student perception on learning (for professional students).

"Normally, I set the scores of junior grades between 60 and 80, and that of senior grades would be 70–90. That's to say, for example, in junior grades, those who are below average will get 60–70, those who are the average will get 70–80 and those above average will get 80–90." (T3)

Interestingly, no amateur students or their parents mentioned the test score. Because to avoid controversies about the unconvincing scoring in assessment, Children's Palace employed other approaches as alternatives, such as awarding stars or merit certificates and developing "a three-level promotion mechanism":

> "We seldom give scores or comments in written form to students in daily teaching and learning. Instead, we employ a three-level promotion mechanism, which we think is more practical for amateur students and more suitable for the teaching and learning of the Children's Palace. The three-level promotion mechanism is a kind of practical assessment system. Starting with the regular class, students can be promoted to art ensemble if they pass the examination, and they would be selected to take part in competitions if outstanding performance is achieved." (T1)

Table 10. Data tabulation of the assessment results after the pilot implementation

Location	Teacher	Student	Singing skills (Psychomotor Domain)	Q&A (Cognitive domain)	Singing Class in Cantonese Operatic Singing (Performance Domain)	Level of Usual Academic Performance	Grades
Children's Palace	T-1	T1-S1	50	55	55	Underachiever	Junior Grades
		T1-S2	51	63.33	58.33	Average	
		T1-S3	88	85	85.33	Top	
		T1-S4	75.83	85	78.33	Underachiever	SeniorGrades
		T1-S5	80	80.33	80	Average	
		T1-S6	85.83	81.67	85	Top	
Cantonese Opera School	T-2	T2-S1	57	55	55	Underachiever	Junior Grades
		T2-S2	75	60	67.5	Average	
		T2-S3	85	61.67	76.67	Top	
		T2-S4	70	71.67	69.17	Underachiever	SeniorGrades
		T2-S5	75.83	75	79.17	Average	
		T2-S6	85.83	86.67	84.17	Top	
	T-3	T3-S1	66.5	81	68.75	Underachiever	Junior Grades
		T3-S2	86.5	70	77.75	Average	
		T3-S3	82.5	73	84.5	Top	
		T3-S4	69.5	54	69.25	Underachiever	SeniorGrades
		T3-S5	75.75	58	74.75	Average	
		T3-S6	89.75	67.25	84.5	Top	
	T-4	T4-S1	55	51.67	55	Underachiever	Junior Grades
		T4-S2	85.5	65.5	80.5	Average	
		T4-S3	89	71.67	88.33	Top	
		T4-S4	66.5	53	68	Underachiever	SeniorGrades
		T4-S5	81.5	58.5	80.5	Average	
		T4-S6	88	66.67	85	Top	

From both the perspectives of teachers and professional students, "test score manipulation" helps to "avoid failure rate" and will not "hurt the student self-esteem" with scores below the pass rate. They attributed the third reason for "test score manipulation" to "affected by human factors", such as bias or preference.

> "Nowadays, the students are very fragile psychologically. I am afraid it might hurt their learning initiative if I give an F to them." (T2)

> "I think the scores that we got were relatively higher than our ability. It might be because the teachers were careful of our self-esteem and might thus lead to the involvement of personal emotions." (T3-S4)
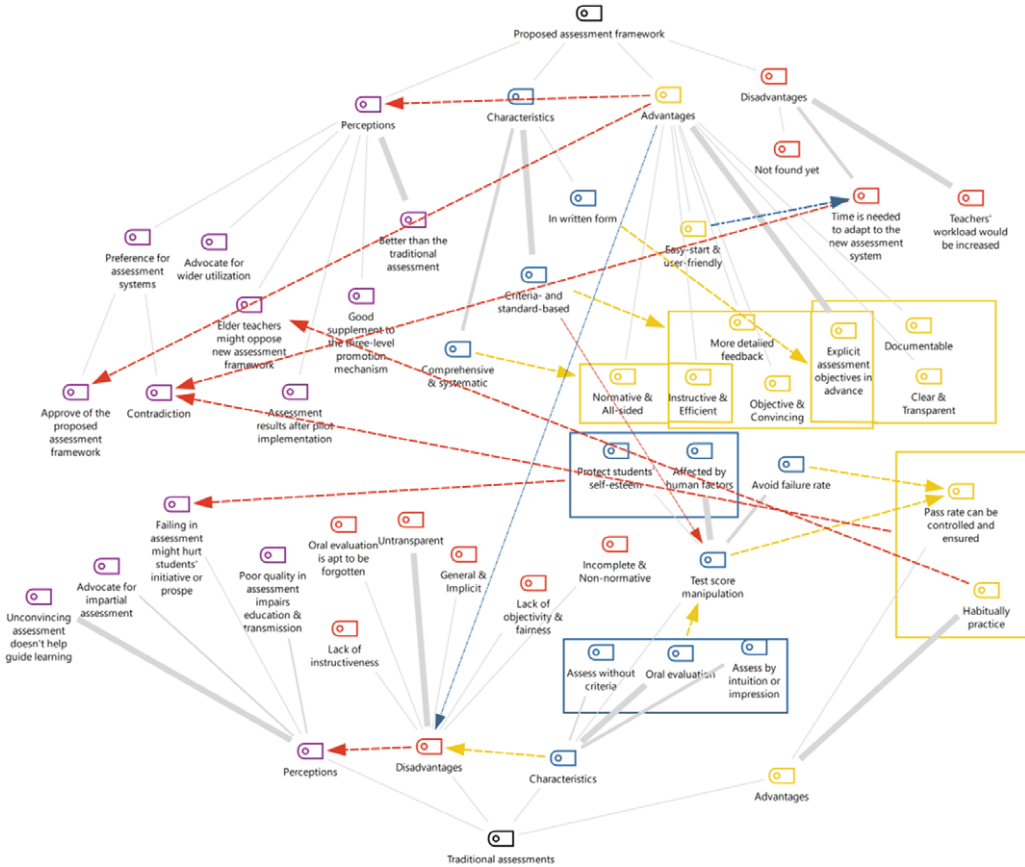
Figure 9. Thematic networks of the semi-structured interview.
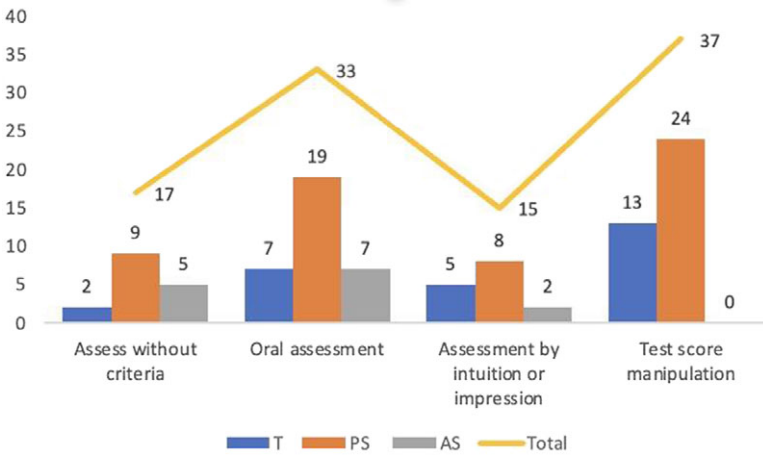
Figure 10. The characteristics of traditional assessment practice.

Table 11. A transcript of professional student academic performance in a semester

No.	Name	Mid-term Grade	Final Grade	Total Grade
1	HJL	85	88	87
2	LXC	92	96	94
3	FLQ	92	95	94
4	YTR	86	88	87
5	ZZY	90	96	94
6	OYSQ	78	84	82
7	YYX	83	86	85
8	HZZ	90	92	91
9	LJM	60	60	60
10	LLD	78	82	80
11	HJJ	90	94	92
12	JZT	88	94	92

> "It is even more overt in the assessment of competitions. The scoring is usually affected by who is whose student and who is whose master. It is a common phenomenon in the Chinese theatrical circle." (T4)

Furthermore, the teachers (mentioned seven times), professional students (mentioned 19 times) and amateur students (mentioned seven times) admitted the prevalence of "oral assessment". It is the second noticeable characteristic of the traditional assessment practice.

> "Unlike your assessment framework, what we received before was a total point and some overall oral comments as assessment. Although the oral comments were helpful, they were too apt to be forgotten and not as clear and orderly as your assessment sheet. (T3-S1)

"Assess without criteria" is another characteristic that cannot be ignored, which might result in some of the disadvantages of the traditional assessment practice."

> "Since there are no unified criteria or standards, some teachers are rigorous, whereas others are loose when scoring. Consequently, some scores are relatively low, whereas some are relatively high, which makes the comparison amongst students unworkable and confuses the students: how exactly did I perform?" (T4)

The last characterises extracted from the interviews was "assess by intuition or impression". It might be because of another characteristic "assess without criteria".

> "Unlike your assessment framework, we don't have unified or explicit assessment criteria and standards in our daily assessment practice. We make assessments out of our expertise and experience. I believe these have already become our intuition." (T2)

> "In my opinion, adjusting some of the scores is very necessary. Sometimes a student might perform not as good as usual in an assessment. But I know my student well, and I know exactly how well he can do actually. Even though he did poorly this time, I was

still inclined to give him a better assessment result according to his level in my mind. I believe this will protect his initiative and self-esteem." (T3)

An internal association exists among the four characteristics of traditional assessment practice: "assess without criteria", "assess by intuition or impression" and "oral assessment" might conspire to lead to "test score manipulation" in the traditional assessment.

"Besides oral comments during the class, we offer total points only, without sub-scores. I think sub-scores come from subentries of assessment criteria. But no one ever bothers to dig into this part. And I think it is hard to do so." (T4)

These four main characteristics of the traditional assessment practice might be the roots of multiple disadvantages of the traditional assessment practice, such as "general & implicit", "lack of objectivity & fairness", "incomplete & non-normative", "untransparent", "oral evaluation is apt to be forgotten" and "lack of instructiveness".

"I haven't seen any criteria and standards in our assessment before. Although the teachers mentioned some requirements in class, I don't know whether those are requirements exactly the criteria in the assessment." (T1-S5)

"Besides a total point, the teachers normally give us some oral evaluations. These comments are helpful but are not as clear or detailed as the results that we got after the pilot implementation of your assessment system." (T3-S1)

Because teachers "avoid failure rate" intentionally and via "test score manipulation" in the traditional assessment practice, "the pass rate can be controlled and ensured", and it even became one of the advantages that they thought.

In the second kind of relationship ("influenced by"), two pairs of relationships were perceived:

1) The disadvantages of the traditional assessment practice, such as "lack of objectivity and fairness", might affect interviewees' perceptions, including "unconvincing assessment doesn't help guide learning", "advocate for impartial assessment" and "poor quality in assessment impairs education & transmission" of Cantonese opera.

   "I think the assessment in Cantonese opera is not as clear and transparent as those of other subjects that we got at school. As parents, it is hard to keep track of our children's learning progress in Cantonese opera. As you know, nowadays children are facing enormous study pressure. A Grade 5 primary student often has to work hard to 9 or 10 at night to finish his homework. So, we don't want to waste our time on the extra-curriculum that we cannot see progress or achievement. I have seen many students quit Cantonese opera after they rise to higher grades in primary school." (T4-S4)

2) The teachers' perception that "failing in assessment might hurt the students' initiative or prospect" might influence their deeds when assessing, which is reflected in the characteristics of traditional assessment practice: "Affected by human factors" and out of the purpose of "protecting students' self-esteem" and "avoiding failure rate" might eventually result in "test score manipulation".

In Theme 2, "proposed assessment framework facilitates teaching & learning", three characteristics were extracted from the interviews, including "criteria- and standard-based", "comprehensive & systematic" and "in written form" (Figure 11).

Four pairs of interactions were detected in the first kind of relationship ("associated with") in Theme 2:

1) As shown in the above schematic, "criteria- and standard-based" is the most prominent characteristics extracted from the interviews, which might contribute to the multiple advantages
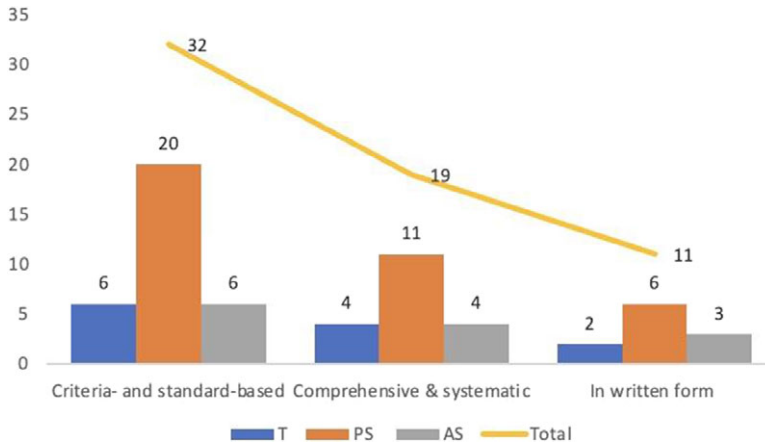
Figure 11. The characteristics of the proposed assessment framework.

of the proposed assessment framework, including "instructive & efficient", "more detailed feed-back", "explicit assessment objectives in advance" and "objective & convincing".

> "My mom and I both think the proposed assessment framework based on "criteria- and standard-based" is a good design. With this assessment sheet, we could get to know what and how we are going to be assessed beforehand. It facilitated me to make better preparation for the assessment. And after the assessment, we received the assessment sheet scored by my teacher. It read not only a total point but also some progress bars corresponding to every criterion. I like these progress bars. They illustrated how well I performed vividly and help to guide my subsequent learning. This kind of feedback is much clearer and more detailed than the assessment that we experienced before." (T1-S1)

The criteria- and standard-based assessment framework was appreciated by the interviewees for it is"easy-start & user-friendly".

> "At first, I worried whether this new assessment framework is difficult to use. But after the pilot implementation, I found it quite easy to start with. And the total point would come out automatically after I clicked the grid of every criterion to judge my students' performance. It is very convenient and efficient." (T1)

2) The second characteristic turned out to be "comprehensive & systematic", which might associate with some of the advantages, such as "normative & all-sided" and "instructive & efficient".
3) The interviewees indicated that the proposed assessment framework also features "in written form". This characteristic might arise in connection with several advantages, including "clear & Transparent", "documentable", "explicit assessment objectives in advance" and "more detailed feedback".

> "This new assessment can provide us with some feedback in written form. I think it is much more instructive than the previous form of assessment that we had. Because we used to get a total point and some oral comments from the teachers, though these comments are helpful, it is hard to remember them all for a long time. Now I don't worry about this problem anymore with the written feedback. And if I keep it, I think I can use it to make comparisons with my previous or future performance. It will help me to keep track of my own study progress." (T2-S6)

4) Although most of the interviews declared the disadvantage was "not found yet", two people expressed their concern that "time is needed to adapt to the new assessment system". It might be due to the characteristics of this proposed assessment framework:

> "I think this assessment framework would greatly normalise the assessment in Cantonese opera. But getting used to employing it might be time-consuming. For example, it took me some time to fix all the criteria for every domain before the pilot implementation, whereas I didn't need to do so before. I just made an assessment holistically by offering a total score and some comments orally. That would be much simpler." (T3)

In the second kind of relationship ("influenced by"), four pairs of relationships were perceived in Theme 2:

1) Regarding "the assessment results after pilot implementation", half interviewees stated that the pilot implementation assessment results coincided with previous ones. The other half of the interviewees noted that it was "lower" than what they gave/received. It might be affected by the characteristics of both the old and new assessments: Since the pilot implementation employed the proposed assessment framework that was "criteria- and standard-based", the behaviour of "test score manipulation" that is common in the traditional assessment practice was likely to be restrained. Consequently, the students gained more "objective" results, which were inevitably inferior to those adjusted via "test score manipulation".

2) "Contradiction", one of the interviewees' options regarding "preference for the assessment system", might be influenced by one of the proposed assessment framework's disadvantages that they were concerned about: "time is needed to adapt to the new assessment framework". Besides, the advantages of the traditional assessment practice might also impact this perception: "habitually practice" and "pass rate can be controlled and ensured" might be the resistance to their embrace of a new assessment framework. By contrast, most interviewees "approve of the proposed assessment framework" when asked about their attitude towards "preference for the assessment system", which might be due to the diverse advantages of the proposed assessment framework.

3) The multiple advantages of the proposed assessment framework might be the main reason for some interviewees' perceptions. These included "more helpful to teaching and learning", "advocate for wider utilisation" of the proposed assessment framework" and "a good supplement to the three-level promotion mechanism" at Children's Palace.

4) Getting to know (and learning to use) a newly developed assessment framework demands extra time and endeavour. The proposed assessment framework requires teachers to specify the criteria in advance and assess specific points corresponding to every criterion rather than give a total score and is, therefore, more detailed. Thus, teachers worried that their "workload might be increased", thereby considering it a disadvantage of the proposed assessment framework. Influenced by this and because the traditional assessment was "habitually practice" that they had employed for years, the "elder teachers might oppose new assessment framework".

In the third kind of relationship ("solved by"), three pairs of relationships were noted between the two themes:

1) The pilot implementation proved that the disadvantages of the traditional assessment practice can be made up by the advantages of the proposed assessment framework:

> "Lack of objectivity & fairness" is the interviewees' primary concern in the traditional assessment practice, which can be mitigated by employing the proposed assessment

framework. Because the interviewees endorsed that the assessment results were "objective & convincing", which benefit from another two advantages of the proposed assessment framework- "explicit assessment objectives in advance" and "more detailed feedback".

"Lack of instructiveness" was deemed one of the traditional assessment's disadvantages, which might be solved if the proposed assessment framework is used since "instructive & efficient" was believed to be one of its advantages. The standards of every subject correspond to the hierarchies of Bloom's Taxonomy in the corresponding domain, and the teachers do not need to bother with it anymore. When applying this assessment framework, what the teachers need to do are: 1) define every criterion according to the assessment purpose or needs before the assessment and 2) click the grid corresponding to every criterion and standard. By doing so, three kinds of assessment results can be produced: 1) automatically calculated total points via algorithms set in Excel in advance, 2) progress bars representing the levels of students' performance and 3) targeted and instructive comments from the teachers according to the above data in the assessment sheets. This practice not only helps the teachers normalise and simplify the assessment process but also provides the teachers with grounds to give their students with overall evaluation and instructive feedback focussed on every detailed criterion.

And the complaints that the traditional assessment practice is "incomplete & non-normative" can be dealt with, for the interviewees indicated that one of the proposed assessment framework's advantages was "normative & all-sided".

Some of the proposed assessment framework's advantages, such as "more detailed feedback", "clear & transparent" and "Explicit assessment objectives in advance", contribute to two of the traditional assessment's disadvantages, "general & implicit" and "oral evaluation is apt to be forgotten".

The "untransparent" assessment results in the traditional assessment practice not only made some of the students confused about how to improve but also hindered the parents from knowing more about their children's progress in study. Both the students and their parents deemed that this problem was solved by the criteria- and standard-based assessment framework, for they endorsed that the assessment in the pilot implementation was "clear & transparent".

2) A tiny minority of interviewees were concerned that "time is needed to adapt to the new assessment framework". This might not be a problem since "easy-start and user-friendly" was considered one of the proposed assessment framework's advantages by most interviewees.

> "Despite the time and endeavour needed to adapt to this new assessment approach, I assume it is unavoidable when embracing a new approach. The diverse advantages of the proposed assessment framework and the benefits that it brings to the assessment of Cantonese opera will make the efforts of introducing and generalising it very worthwhile." (T1)

3) Since the above problems were solved respectively, the negative perceptions from the interviewees were also alleviated, including "poor quality in assessment impairs education & transmission", "unconvincing assessment doesn't help guide learning" and "elder teachers might oppose new assessment framework".

Discussion

The assessment of Cantonese operatic singing is still underdeveloped. The semi-structured interviews revealed that the traditional assessment practice in Cantonese operatic singing featured four characteristics, including "assess without criteria", "test score manipulation", "oral assessment" and "assess by intuition or impression". Multiple disadvantages were derived from these

characteristics, such as "lack of objectivity and fairness", "lack of instructiveness" and so on. These characteristics and disadvantages considerably discredited the authority and authenticity of assessment in this field. Consequently, proposing a criteria- and standard-based assessment framework might be a good attempt to regulate assessment practice in daily teaching and learning and restrain human factors in assessment, thereby providing convincing assessment. Improving the educational assessment of this traditional genre might facilitate helping Cantonese operatic singing thrive in an increasingly accountability-driven educational environment and assessment-oriented world.

Conclusion

The feedback from the semi-structured interviews turned out to be positive, affirming that the advantages of the proposed assessment framework far outweigh its disadvantages and contribute to making up for the deficiencies in the traditional assessment practice. Gardner (2012) indicated that: "assessment in education must, first and foremost, serve the purpose of supporting learning" (p. 9). Most of the interviewees endorsed the efficacy of the proposed assessment framework. Its characteristics, such as "criteria- and standard-based", "comprehensive & systematic" and "in written form", produces diverse advantages of this new approach, including "objective & convincing", "instructive & efficient", "normative & all-sided", "easy-start & user-friendly", "more detailed feedback", "explicit assessment objectives in advance" and "documentable". These advantages are imperative to any newly proposed assessment framework, especially from an educational perspective. Both professional and amateur students, as well as their parents, stated that the new assessment framework is more instructive than traditional methods, which achieves the overarching purpose of assessment and facilitates the teaching and learning of Cantonese operatic singing across multiple aspects.

Implications

Based on a developed assessment framework for Cantonese operatic singing, this study has filled a gap in the related literature. Considering the statement of limitations of this research and previous discussions, future research might concentrate on the following four aspects: 1) in-depth educational research into Cantonese operatic singing assessments as breakthrough points; 2) research into administration or policy strategies surrounding the education and transmission of Cantonese operatic singing based on the data derived from assessment; 3) the transmission and development of the traditional genre through educational influence within contemporary society and 4) further research into music education assessment.

References

AIRASIAN, W. & MIRANDA, H. (2002). The role of assessment in the revised taxonomy. *Theory into Practice*, **41**(4), 249–254. https://doi.org/10.1207/s15430421tip4104_8

ANDERSON, L. W. & KRATHWOHL, D. R. (2001). *A taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. London: Longman.

ANDERSON, L. W. & SOSNIAK, L. A. (1994). *Bloom's Taxonomy: A Forty-year Retrospective*. Chicago: NSSE.

BLANTON, H. & JACCARD, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, **61**(1), 27. https://doi.apa.org/doi/10.1037/0003-066X.61.1.27

BLOOM, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. London: Longman.

CHAN, S. Y. (1999). 香港粵劇導論 [*Introduction to Cantonese Opera in Hong Kong*]. Hong Kong: The Chinese University Press.

CLARKE, V. & BRAUN, V. (2014). Thematic Analysis. In T. TEO (Ed.), *Encyclopedia of Critical Psychology* (pp. 1947–1952). New York: Springer.

COHEN, L., MANION, L. & MORRISON, K. (2018). *Research Methods in Education*. New York: Routledge.

DILLMAN, D. A., PHELPS, G., TORTORA, R., SWIFT, K., KOHRELL, J., BERCK, J. & MESSER, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR), and the Internet. *Social Science Research*, **38**(1), 1–18. https://doi.org/10.1016/j.ssresearch.2008.03.007

GARDNER, J. (2012). *Assessment and Learning* (2nd ed.). London: SAGE.

GEISERT, P. (1972). *The Dimensions of Measurement of the Affective Domain (ED069663)*. ERIC. https://files.eric.ed.gov/fulltext/ED069663.pdf

GREENWALD, N. B. A. & BANAJI, M. R. (2003). Understanding and using the implicit association test. *Journal of Personality and Social Psychology*, **85**(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

GRONLUND, N. E. & BROOKHART, S. M. (2009). *Writing Instructional Objectives* (8th ed.). Upper Saddle River, NJ: Pearson Education.

HALL, R. A. (2011). Affective assessment: The missing piece of the educational reform puzzle. *Delta Kappa Gamma Bulletin*, **77**(2), 7–10. https://www.semanticscholar.org/paper/Affective-Assessment%3A-The-Missing-Piece-of-the-Hall/f392ab6ddaf9c4680eea2b1bc2b5ee82bf1d5c0f

HAUENSTEIN, A. D. (1998). *A Conceptual Framework for Educational Objectives: A Holistic Approach to Traditional Taxonomies*. Lanham, Maryland: University Press of America.

HERTZUM, M., BANSLER, J. P., HAVN, E., & SIMONSEN, J. (2012). Pilot implementation: Learning from field tests in IS development. *Communications of the Association for Information Systems*, **30**(1), 313–328. https://doi.org/10.17705/1CAIS.03020

INTRINSIC MOTIVATION INVENTORY (n.d.). *Center for Self-Determination Theory*. Retrieved from https://selfdeterminationtheory.org/intrinsic-motivation-inventory/

KRATHWOHL, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice*, **41**(4), 212–218.

KRATHWOHL, D. R., BLOOM, B. S. & MASIA, B. B. (1964). *Taxonomy of Educational Objectives, the Classification of Educational Goals. Handbook II: Affective Domain*. New York: David McKay Co., Inc.

LEUNG, B. W. (2015). Transmission of Cantonese opera in the conservatory tradition: two case studies in South China and Hong Kong. *Music Education Research*, **17**(4), 480–498. https://doi.org/10.1080/14613808.2014.986081

LEUNG, P. K. (1982). 粵劇論 [General Theory of Cantonese Opera Research]. Hong Kong: Cantonese Opera Research Project Press.

LUO, Y., & LEUNG, B. W. (2022). Proposing an assessment framework for Cantonese operatic singing after reviewing the current practices in Hong Kong and Guangdong, China. *Music Education Research*, **25**(1), 1–16. https://doi.org/10.1080/14613808.2022.2156490

MCCOACH, D. B., GABLE, R. K., & MADURA, J. P. (2013). *Instrument Development in the Affective Domain* (3rd ed.). New York: Springer.

OAKLAND, T. (1997). Affective assessment. *Psicologia Escolar e Educacional*, **1**(2–3), 11–21. https://www.scielo.br/j/pee/a/NGRYV73vCJgkTfPPd77mqCn/abstract/?lang=en

PAULHUS, D. L. (1991). Measurement and control of response bias. In J. P. ROBINSON, P. R. SHAVER, & L. S. WRIGHTSMAN (eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 17–59). San Diego: Academic Press.

PAULHUS, D. L. & VAZIRE, S. (2007). The self-report method. In R. FRALEY & R. F. KRUEGER (eds.), *Handbook of research methods in personality psychology* (1st ed., pp. 224–239). New York: Guilford Publications.

REID, N. (2006). Thoughts on attitude measurement. *Research in Science & Technological Education*, **24**(1), 3–27. https://doi.org/10.1080/02635140500485332

WEISBERG, H., WEISBERG, H. F., KROSNICK, J. A. & BOWEN, B. D. (1996). *An Introduction to Survey Research, Polling, and Data Analysis* (3rd ed.). Thousand Oak, California: SAGE Publications.

WILSON, T., LINDSEY, S. & SCHOOLER, T. Y. (2000). A model of dual attitudes. *Psychological Review*, **107**(1), 101–126. https://doi.apa.org/doi/10.1037/0033-295X.107.1.101

ZAI, L., FAN, D. Z., CAO, J., & CHOU, S. Y. (2018). 从传统到现代:学徒制背景下 师徒关系的转变研究 [From Tradition to Modernity: A Study of the Transformation of the Teacher-Apprentice Relationship in the Context of Apprenticeship]. Jiangsu Education Research (Z6), 32–35. https://doi.org/ 10.13696/j.cnki.jer1673-9094.2018.z6.008

ZHANG, K.-Y., & CERDIN, J-L. (2020). The Chinese apprenticeship model: the spirit of craftsmanship. In J.-L. Cerdin & J.-M. Peretti (eds.), *The success of apprenticeships: Views of stakeholders on training and learning*, volume 3 (pp. 187–192). New York: Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119694793