INTERNATIONAL JOURNAL OF SOCIAL SCIENCE AND BUSINESS MANAGEMENT (IJSSBM)
(e-ISSN: 3009-0091) (Vol.02, ISSUE 01,241-257)

MUSIC EDUCATION ASSESSMENTS IN CHINA: BIBLIOMETRIC ANALYSIS

Zhang Yunchao[1], Diyana Kamarudin[2]

[1,2] Faculty of Education and Liberal Studies, City University Malaysia,46100 Petaling Jaya, Selangor, Malaysia
[1] Faculty Vocal and Opera Studies, Sichuan Conservatory Of Music, No.6 Xingsheng Road, Chengdu, China

ABSTRACT-Music assessment is critical to delivering thoughtful, frequent music instruction. The assessment provides information to both instructors and students regarding student achievement and serves as a guide for instructional practices. In China, the teaching system is affected by an exam-oriented habitus, which has resulted in the long-term use of summative assessment to evaluate learning outcomes in the music classroom. Traditional assessment methods have the disadvantage of needing more real-world context. Students answer questions one by one without the need to apply long-term critical reasoning skills. Peer assessment is related to both the processes and the outcome. It is a technique where students assess their performance. Instructors must relinquish their dependence on traditional metrics of technique usage to accommodate this new concept. Peers evaluate student work on projects, assignments, and reports. Innovative teaching techniques and assessments are not only the exterior expression of innovation in music education but also the focal point of this innovation. When it comes to adapting to the demands of the new millennium, music education assessment is a strategic decision for the twenty-first century. The purpose of this study is to review the music assessment methods in Chinese higher education through a bibliometric analysis.

ARTICLE HISTORY
Received: 01/10/2023
Revised: 20/11/2023
Accepted: 1/12/2023

KEYWORDS
Peer Assessment, Traditional Assessment, Higher Education, Music Assessment

INTRODUCTION

Exams are one of many ways to assess. Student development is tracked via assessments, which impact the learning process at all levels (Ravitch, 2020). Assessment typically influences student learning in higher education because it reflects institutional goals. Because of this, it has a considerable backwash effect on teaching and learning (Boud & Falchikov, 2007). Self- and peer assessment has recently gained popularity in higher education to improve learning outcomes (Wanner & Palmer, 2018). In other words, self-and peer assessment helps students learn to evaluate their own and others' work (Carless, 2015).

Ballantyne, Hughes, and Mylonas (2002) state that examining students' or peers' work helps them become more motivated, learn and assimilate topic content, and comprehend the grading process. Self- and peer assessment also encourages students to monitor and evaluate their and their classmates' progress. Students develop lifelong assessment skills by assessing their own and others' work. Students also develop self-directed and independent learning strategies based on their evaluations (Janes, 2007).

Peer evaluation is an evaluation method. Peer evaluation affects both the process and the result. It is a method where students evaluate their work (Gyamli et al., 2022). To accommodate this new approach, instructors must abandon conventional method utilisation measurements. The Turkish Ministry of Education says peers evaluate student projects, assignments, and reports. Students consider each other's work on projects, tasks, and reports. Students must determine their peers using the discipline's standards (Willey, 2018)—another modern evaluative approach. Students will be more involved in the assessment process and increase their knowledge of the subject matter. This enhances teaching techniques and provides students with more hands-on experience. Students feel more responsible when they make remarks. Students need to develop self-awareness and self-awareness of Achieving knowledge, motivation, and trust while improving communication and interpersonal skills and overall well-being may be possible. While many instructors believe this is a valuable formative assessment tool, it has certain drawbacks.

To succeed, the instructor's consent is required. Instructors question the method's reliability compared to traditional teacher evaluation. There is no proof that well-designed observed peer evaluation yields trustworthy peer evaluation (Reinhold, 2016).

PROBLEM STATEMENT

Traditional assessment is a notion that many students may see as an arbitrary evaluation procedure carried out by their instructors (Choriev et al., 2022). Conventional assessment methods have the disadvantage of needing more real-world context. Students answer questions one by one without the need to apply long-term critical reasoning skills. They also need more chances to demonstrate their reasoning skills despite a lack of knowledge about a question's specific subject matter. (Wanner & Palmer, 2018). Alternative assessment methods allow students to apply their skills and knowledge within a context resembling problem-solving and knowledge application in most jobs or daily tasks. Due to recent pedagogical criticism, traditional evaluation methods were replaced by an assessment process in which students and teachers are responsible for judging student performance and achievement, rather than being limited to teachers with students playing an inactive role (Wanner & Palmer, 2018). According to Gaytan (2002), types of evaluation restricted to leading questions only sometimes result in student learning; therefore, to accomplish targeted learning, instructors must keep this in mind while establishing the different criteria, goals, and intended results in the assessment procedures.

Today, there are advancements in assessment techniques, with the shift from summative to formative assessment as the primary goal of the process (Pecheritsa, 2022). These innovations include considering alternatives, which necessitates challenging the learning process and integrating learning and evaluation activities rather than relying just on routine testing apps to achieve success. Alternative assessment allows students to demonstrate what they can achieve; that is, pupils are assessed on what they integrate and generate rather than on what they can recollect or recall (Coombe et al., 2007). As a result, a great variety of unique techniques for assessment have been proposed, all of which aim to improve the integration of learning and assessment by increasing the level of participation of students in the assessment tasks themselves (Sluijsmans et al., 2003). According to Saito (2008), peer assessment promotes reflective learning by allowing students to see others' performances and become aware of the criteria for evaluation. Peer assessment elicits positive reactions from students despite some expressing concerns and worries. It promotes the development of self-awareness by highlighting the differences between one's own and other's perceptions, as well as facilitating further learning and accepting responsibility for it. Additionally, concentrating on peers' strengths and faults may help students learn more effectively, improve their critical thinking, and develop their sense of self-direction.

As a result, a shift away from traditional evaluation methods has occurred. In response to a growing awareness of the impact of testing on curriculum and instruction, educational reformers began to use alternative assessments (Dietel et al., 1991). Similarly, Reeves argued that alternative assessment methodologies challenge standard assessment, often called testing (Reeves, 2000). So, the researcher identified the research problem as “To examine the teacher’s view of Peer Assessment as a Student Evaluation Technique.”

LITERATURE REVIEW Music Education

Music is something that brings us all together. Music is something that everyone can engage in and appreciate, regardless of their culture, social status, or political affiliation (Thompson et al., 2022). Music programs in public schools are critical to continuing the heritage of music-making in our society, and for many students, they serve as the beginning of their musical careers.

Evaluation

Excellent and bad assessments can be characterised, and good and bad evaluators may be differentiated. To some extent, each formal evaluation process that relies on an "expert" assessor is predicated on this notion (Thompson & Williamon, 2003). It would be a mistake to take the idea of a competent, impartial examiner for granted. Research has shown a worrying degree of variability and subjectivity in music performance quality evaluations (Thompson & Williamon, 2003; Waddell &

Williamon, 2017a for reviews) despite the critical role such assessments play in the development and careers of musicians (see McPherson & Schubert, 2004; Griffiths, 2008 for reviews) (Russell, 2015; Kopiez et al., 2017). Expertise is unnecessary for an evaluator to provide consistent and trustworthy assessments. However, some prior research has cast doubt on this assumption (e.g., Fiske, 1975, 1977; Winter, 1993).

Within the same session, there was a wide range of quality in the feedback provided by music instructors, according to a meta-analysis of 86 papers on the subject (Duke, 1999). This isn't to suggest that being able to evaluate music isn't a vital talent or that there isn't a difference between being a good performer and a good judge; instead, it is to underline that musical performance competence isn't always a predictor of judging success. Persons who enter the field of instrumental music teaching (and, by extension, music examination or competition judging) are only sometimes those who have extensive training in assessment but have shown substantial competence in the particular area on which they are giving judgment, namely performance.

Evaluation as Music Performance

Although assessing performance may be considered a distinct competency to be honed, there is merit in seeing it as a performance in and of itself. Expertise is required, just like the musical performance it aims to assess. It happens in certain places and often involves collaborating with people who may or may not have a particular expertise. As in public contests, this might happen in front of an audience that may be judgmental of the results. Consequences arise for the evaluated and evaluative performer regarding their professional status, reputation, and potential for future employment. Furthermore, it is a procedure that occurs in a predetermined order and lasts for a certain length, with separate pre- and post-performance intervals that frequently restrict or forbid opportunities for stop, repetition, or reflection. Taking a performance-based approach to assessment prompts us to rethink how to handle the process best. The evaluation method may be evaluated for its quality and usefulness, making evaluation not merely a tool for summarizing, diagnosing, and developing performance.

In this perspective, meta-assessment takes the form of figuring out how to provide a formative assessment on formative assessment since the expertise required to carry out a successful evaluation has become a sort of assessment itself. The seven principles of evaluation described above (Nicol & Macfarlane-Dick, 2006) may be applied to the assessment evaluation itself if one views evaluation as a performance.

This theory emphasises the actual application of the skill being honed, reinforcing the centrality of self-regulated learning. This is a central tenet of experiential learning, which argues that students learn best when they construct their knowledge via active participation in and reflection on various meaningful situations (Kolb & Kolb, 2005). By combining the ideas that assessment is both a learned and practised talent, we may use the framework provided by current performance training techniques that incorporate experiential learning to develop novel evaluation training and research approaches.

Dress rehearsals are a tried-and-true method for musicians to practice for actual performances by removing the audience and focusing on the technical parts of the show. While this may include checking the timing, costume, and operation of any electrical or mechanical parts, it also allows the performers to test their technical, physical, and mental preparedness for the performance. Notably, the dress rehearsal provides an opportunity to address the increased physiological arousal that accompanies the performance, and that may have a maladaptive effect on results if the performer misinterprets it as the expression of performance anxiety (Kenny, 2011; Nieuwenhuys & Oudejans, 2012; Endo et al., 2014). This is true not just of the time spent on stage but also of the time spent backstage before the performance when performance-related physiological arousal is at its maximum (Williamon et al., 2014; Chanwimalueang et al., 2017). The stress of an actual performance may be simulated by filming these sessions, which has been shown in studies to make student performers nervous (Daniel, 2001).

Assessment has long served as a valuable tool for learning by doing in the classroom. Indeed, there has been a rise in the use of self- and peer-assessment as a component of the learning process in higher education, with one meta-analysis showing a pattern of good correlations between peer- and professor evaluations provided that global criteria are being applied in both (Falchikov & Goldfinch, 2000).

Classes focused on music pedagogy will examine various instructional paradigms and strategies for providing constructive criticism. The teacher may perform in-class mock lessons or videotape them for later evaluation, but this involves finding students willing to participate in the teaching experiment. In the masterclass or studio class, the experienced musician works with one or more musicians in front of an audience (in the case of the masterclass) or other students, representing a more conventional method of instruction (i.e., the studio class; Gaunt, 2017). This framework may include a panel of specialists, independent study, student-led instruction, or, most importantly, peer assessment (Long et al., 2012). The master/studio class provides a platform where musicians can test and develop their skills of attentive listening and viewing, performance diagnosis, and public performance. For teachers, the master/studio class provides an opportunity to gain exposure as a master teacher, to reach and recruit new students, and to hone their evaluative skills (Hanken, 2008, 2010; Taylor, 2010; Long et al., 2012; Haddon, 2014; Gaunt, 2017).

The topic and duration prescribed by the instructor determine whether a masterclass or studio session offers a unique opportunity to analyse the quality of feedback delivery. In other words, giving an assessment is more likely to encourage contemplation of the performative skill than the evaluative. Research on assessing one's own and others' performances using video recordings reveals the importance of comments centred on the performance itself (e.g., Bergee, 1993, 1997; Johnston, 1993; Robinson, 1993). According to a preliminary questionnaire, less than half of the undergraduate music students surveyed by Daniel (2001) at an Australian institution who participated in his study of videoassisted self-assessment evaluated audio or video recordings of their performance frequently.

Having students provide feedback to one another as part of their training has been the subject of several studies, many of which have evaluated live pilot projects. Throughout many years, Hunter and Russ (1996) collaborated with an Irish institution to design and oversee a seminar on peer evaluation. Students were instructed in the university's evaluation methods and then formed into panels consisting of students with varying levels of instrumental expertise, a self-elected leader, and a supporting member of staff who had first instructed the students in the procedures. Several extra-performance biases and complexities were expressed directly in post-evaluation conversations among the students, as shown by a later study. These included the students' realisation that it was socially and emotionally challenging to award a low grade despite a terrible performance, that assessors playing the same instrument as the performer were harsher in their critique than those without the specialised skill, and that marks provided frequently mirrored pre-exhibit prejudices.

Beginning with a pilot in composition and eventually including performance, business, technology, and theory, Searby and Ewers (1997) studied the implementation of a peer evaluation method throughout a university's music department in the United Kingdom. Students in both groups selected the standards for evaluation, had some practice with reviewing work from the previous year, assessed each other's work with the lecturer serving as a moderator, and had their written comments count for 20% of their final grade. Instead of using previously recorded material, the peer assessment method was tested using the musical performances of a different year group. After further discussion with the students, it became clear that the groups' ongoing process of negotiating new sets of evaluative criteria was an integral part of the students' internalising the evaluation process and developing critical thinking skills in producing their work for assessment. This process evaluation also showed that students valued and appreciated peer input as a means of improvement. Unfortunately, peer evaluation did not significantly lessen the burden of faculty members engaged in evaluating students.

Bergee and Cecconi-Roberts (2002) conducted an experiment in which groups of three to five undergraduate music majors performed for one another in four video-recorded sessions, then reviewed and discussed the performance footage while completing self- and peer-assessments using fixed rubrics in light of two studies showing students' inconsistency in their self- and peer-assessment abilities compared with faculty-generated scores (Bergee, 1993, 1997). The jury's students and examiners listened to recordings of their final performances and rated them on a scale from 1 to 10. The capacity to self-evaluate did not vary significantly by year or performance level, and the correlations between students' self-evaluations and faculty evaluations were only slightly more significant in the experimental group than in the control group. Nonetheless, there was still a considerable degree of variation in the evaluations, particularly about assessments of tone and interpretation. The authors suggest that the interventions had not fully engaged with the social and environmental complexities of performance self-

assessment because a subsequent experiment that included a more significant discussion of the evaluative criteria and their application to two sample scores also showed moderate to no effect of the treatment on the alignment of self- and peer-assessments with faculty assessments.

Using a segmented grading method, Daniel (2004) had 36 students in weekly performance seminars evaluate their peers' performances and offer brief remarks and in-depth scores. Students who were initially overly restrained in their critical judgements showed substantial improvement after the sessions, as demonstrated by the results of reflective surveys.

Blom and Poole (2004) surveyed 16 advanced music majors at an Australian institution about how they felt about judging the performances of first-year students. Students were asked to grade recorded performances of their peers using the same criteria employed by staff, provide written critiques to be read by the performers, assign grades, and provide a self-reflective commentary on the process after completing self-assessment tasks and paired peer-assessment critiques in the first year. It was difficult for students to evaluate their peers in a range of instrumental areas, to provide constructive criticism when they were already acquainted with the performance, to use a consistent set of criteria, and to feel like they had the "power" to do so. Hunter and Russ (1996) showed that students benefited from the activity by learning to judge their performance and increasing self-assurance. Following up on Hunter and Ross's use of student-selected evaluation criteria, additional studies found that when evaluating the quality of a rehearsal, students were more concerned with "soft" skills like self-awareness and social awareness, while when assessing the quality of a performance, they were more concerned with "hard" skills like technique, analysis, and musicianship (Blom & Encarnacao, 2012).

When teaching a course on famous music creation at an Australian university, Lebler (2007) outlined how students established a "masterless studio" in which they took charge of their learning techniques, objectives, and outputs while collaborating with their peers. Over a semester, students provided comments totalling over 180,000 words on 292 recorded tracks using a systematised peer assessment process that included sharing recordings and written discussions on a course website. Even though no training or emphasis was placed on providing feedback effectively, course conveners checked whether the input corresponded to excellent norms of constructive criticism, calling out instances of unduly authoritative tone or lack of relevant information.

The undergraduate voice students at an Australian institution were the subjects of Latukefu's (2010) investigation of a scaffolded peer assessment system. Before rolling out the program to a whole class, student focus groups developed the evaluation rubric and procedures based on the framework Searby and Ewers (1997) outlined. Panels of three students conducted peer assessments after the criteria were distributed and discussed in a seminar on modern performance practice. Students acknowledged the value of peer review in enhancing their capacity for self-reflection and skill development as future professionals by responding positively to an open-ended survey. Awkwardness and social factors, they said, inhibit objective discussions of performance and assessment when conducted with friends and peers.

The Norwegian Academy of Music's Center for Excellence in Music Performance Education has adopted group instruction and learning among students as a "primary instrument study" (Hanken, 2016). Several strategies were used, all variations on the classic studio class in which students discuss their work and get constructive criticism from an instructor. One method involved using the Critical Response Process developed by Lerman and Borstel (2003). This process involves the following steps: an initial discussion of what aspects of the performance are meaningful; the performer asking questions on which they would like feedback; the evaluators asking neutral questions of the performer; and the evaluators asking permission to give opinions on specific aspects of the performance; delivering those opinions only if requested. Based on the results of this research, the fourth step of the approach is unnecessary after the performer has arrived at the necessary conclusions via the discourse. To counteract the isolation that might be intrinsic to music education due to the nature of working practices, Hanken also emphasised the value of peer learning in CPD for music instructors via seminars and discussions.

Mitchell and Benedict (2017) used peer assessment in the context of university auditions in Australia. To confront audio/video interaction issues inherent to music performance evaluation directly, we did not have students provide ratings in authentic grading scenarios. Instead, we had them rate live

performances with or without a blinding screen in front of the stage and recorded performances in audioonly, visual-only, and audio-visual scenarios. Judging performances based on audio alone gave the student judges greater confidence and made them think about the importance of presentation in their performances.

Last but not least, Dotger et al. (2018) used techniques in medical education to educate doctors, focusing on a particular way of feedback delivery among music educators. Similar to how a medical student might interact with a dummy patient, this study involved 13 student music educators interacting with a dummy parent who had been instructed to ask questions about the teacher's explanation for her daughter's failure at a recent (fictitious) audition, the reliability of the evaluation, and whether or not her daughter possessed "the look" (i.e., whether she conformed to the presumed stereotypes of performer appearance). Participants' replies were widely divergent since they had yet to be briefed on how to handle the conversation. Several addressed the parents' worries, shared their own experiences, and offered concrete suggestions for improvement in one conversation.

When compared, there are several commonalities between these methods. Each one embraced experiential learning by putting students in charge of the assessment process and, in many instances, the outcomes. When asked about the results, both students and teachers were generally supportive. While it may seem like a no-brainer to give students practice evaluating their peers, some of the studies acknowledge that many students still needed help to handle the stresses of real-world assessment scenarios and point out the workload costs of offering such training. Competitions, auditions, examinations, and masterclasses where the students will be called upon to make consequential judgments stand in stark contrast to artificially contrived evaluations amid familiar peers and surroundings. However, allowing students (or researchers) access to authentic assessment scenarios strips them of agency. It could compromise the evaluation quality, mainly if the students or researchers need to be more experienced.

Therefore, a means must be found to simulate the complexities of a natural or fake assessment while allowing for complete manipulation of the evaluation's input and environment. Dotger et al. (2018), who conducted the fake-parent research, characterise their method as a sort of simulation, distinguishing it from a role-playing exercise by informing participants that the mock parent would never deviate from their character and that the interaction could not be paused or redone. One method that already considers the idea of simulation is using IVEs (IVEs).

RESEARCH METHODOLOGY

The bibliometrics research approach is a tool for creating a "scientific road map" that explores the interconnections between different academic topics, researchers, and published works (Zupic & Carter, 2015). Scholars are very interested in this technique since it combines classification and visualisation to create a map of the structure of the scientific area (Boyack & Klavans, 2014; Van Eck & Waltman, 2014). It is among the most potent tools in literature reviews using scientific metric database networks. This research, which uses a bibliometric method, concludes a publishing cycle on peer assessment, in this case, concerning music education in China.

Data Collection

For this search, the researcher compiled a set of keywords from several sources, including dictionaries, encyclopedias, and other studies, to use as the primary keywords and three supporting terms. 1) peer assessment, 2) Chinese higher education institutions, and 3) music education. The search terms for peer review and related topics were mainly derived from relevant articles. The search phrase for Music education includes "Music assessment" or "music evaluation." In contrast, the search string for higher education institutions contains "higher education institution" OR "university*" OR "college*" OR "institution of higher learning" OR "faculty." The "subject" field is utilised, and the publications' title, abstract, and keywords are searched. The scope of the study was narrowed from "all fields" to just the papers in the field, as mentioned above, to avoid any accidental appearances in content that is entirely out of place. There was no linguistic barrier or other limits placed on the retrieval process. Only the Chinese geographical area can serve as a barrier. All of the articles that were found were dated between

1999 and 2022. The date of the database search was April 2022. The central databases used for the search were Scopus and ScienceDirect. The researcher modified a small number of databases to extract more valuable articles. A first search yielded 2,783 results.

Sample

The database search has returned a total of 2,783 publications. Besides that, researchers have conducted a backwards-forward search from the reference lists of other most relevant and good-quality reviews of music assessment studies previously done. Then, the researcher filtered the publications to only full-text access and written in English only. This was easier since Scopus and ScienceDirect databases mainly consisted of English articles. Finally, the researcher obtained a total of 98 publications.

ANALYSIS

The analysis was done using VOS software, and the following sections describe the analysis and findings of the study. Since the limitations of articles and the interconnection of the studies, the results and analysis were limited to some extent.

Citation Analysis

The most reviewed concepts in music assessment of higher education in China are identified through the bibliometric document citation analysis. In total, 2,783 references from 98 articles were extracted from articles related to music assessment in higher education in China. The analysis led to 98 top articles by selecting a cut-off of articles with one or more citations.

Co-occurrence Analysis

The keyword co-occurrence analysis was based on the author's keyword analysis. The threshold of at least ten occurrences was set (Ahmad et al., 2021). From the 357 keywords, 47 keywords were processed. Keywords with higher frequency are a statement of assessment methods in music education. The most increased word occurrence was Music Education (18), followed by Students (13) and Peer Assessment (9). The table presents the top 10 highest frequencies of the co-occurrence of keyword analysis.

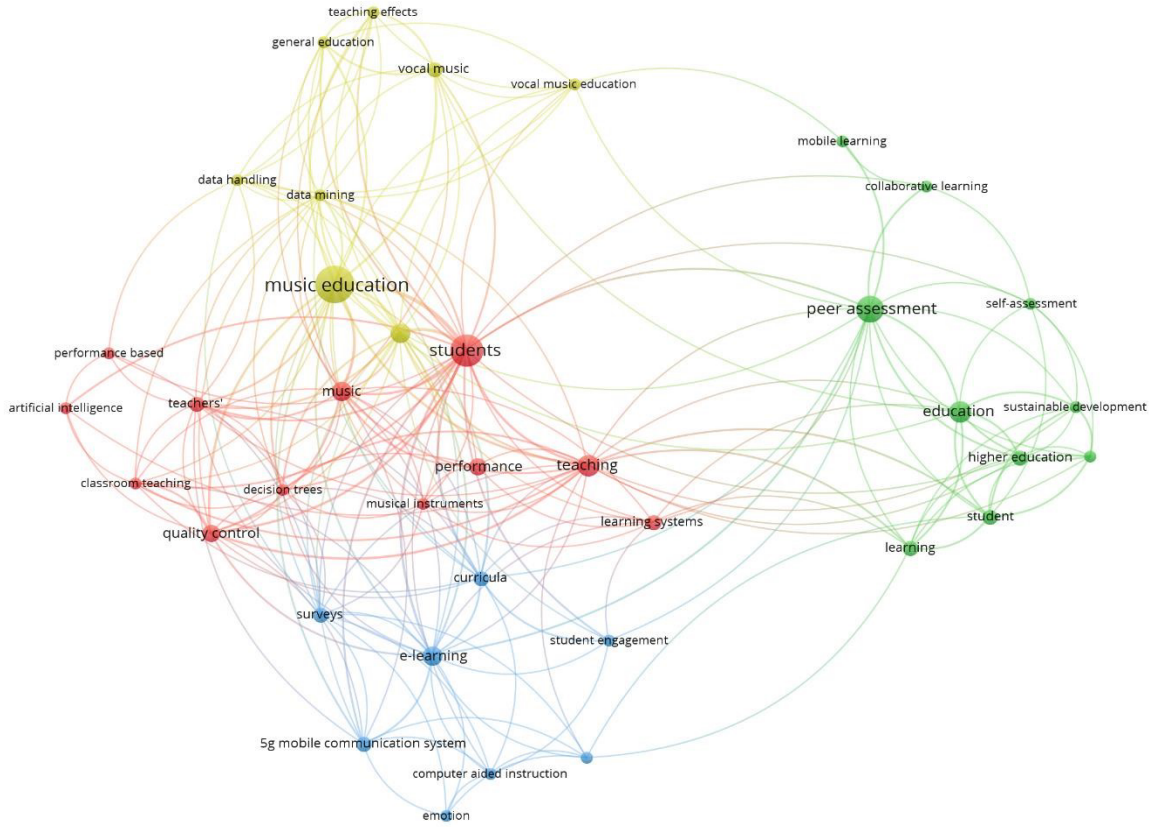

Table 1. Top 10 keywords in the co-occurrence of keywords analysis.

Rank	Keyword	Co-occurrence	Link Strength
1	Music Education	18	61
2	Students	13	63
3	Peer Assessment	9	25
4	Performance	4	9
5	Quality Control	4	24
6	Collaborative Learning	2	6
7	Self-Assessment	2	8
8	Student Engagement	2	10
9	Performance-based	2	5
10	Computer-aided Instruction	2	9

Cluster Analysis

This research could not perform co-citation analysis due to limited access to full-text articles. While 2,783 references were extracted from articles related to music assessment in higher education in China, obtaining full-text access to all these articles proved challenging, making it infeasible to identify shared citation relationships accurately. Instead, the study opted for co-occurrence analysis, which explored the frequency and associations of specific concepts within the available references, providing valuable insights into the prevalence of critical themes in music assessment in Chinese higher education.

Table 2. Clusters and Keywords

Cluster	Keywords	Theme

Red	Quality Control, Performance-based, Performance, Learning Systems, Music Instruments, Classroom teaching	Music Assessment Parameters
Green	Collaborative Learning, Peer Assessment, Selfassessment, Mobile Learning, Sustainable development	Assessment methods
Blue	Computer-aided instruction, e-learning, Knowledge construction, Student Engagement	Music Assessment Innovation
Yellow	Music education, teaching effect, Vocal music, education computing	Music Education

Cluster 1 (red) consists of 12 keywords. This cluster represents the theme "Music Assessment Parameters. " The main keywords are quality control, Performance-based, and learning systems. These keywords are relatable in the context of adapting successful assessment methods in music classrooms in universities. There is a trend in assessing students through performance-based evaluation methods in music classrooms (Guo & Xu, 2022).

Cluster 2 (green) presents ten keywords. It is labelled as "Assessment methods." Representative keywords include Collaborative learning, Peer assessment and Self-assessment. This cluster is related to assessment methods instructors apply to evaluate the students. There is a trend of assessing the students through peer and self-assessment methods instead of following traditional exams and another type of evaluation (Hu et al., 2021).

Cluster 3 (blue) comprises nine keywords with computer-aided instruction, e-learning, and student engagement. This cluster is labelled as "Music Assessment Innovation." Recent technologies and innovative support have greatly influenced assessment methods to adapt a successful student evaluation in recent history. (Sun, 2022).

With eight keywords, cluster 4 (yellow) dealt with “Music Education." The main keywords are Music education, teaching effect, Vocal music, and education computing. Chinese music education has been dramatically influenced by the teaching practices and other evaluation methods used through technology and other trends (Liu et al., 2021).

DISCUSSION

Cluster 1 consists of 12 keywords, with the main keywords being "Quality control," "Performance-based," and "Learning systems." The theme represents the focus on establishing effective assessment parameters in music education. Researchers have observed a trend in music classrooms towards evaluating students through performance-based evaluation methods (Guo & Xu, 2022). These assessment parameters are crucial as they determine the quality of music education and ensure students' development in both performance skills and learning outcomes. Previous studies have highlighted the significance of well-defined assessment criteria and rubrics to maintain quality control in music education (e.g., Wilson & Dunlea, 2020). Educators and administrators need to pay attention to this cluster as it provides insights into improving the assessment frameworks in music classrooms, ultimately enhancing the overall learning

Cluster 2 comprises ten keywords, including "Collaborative learning," "Peer assessment," and "Self-assessment." The theme revolves around various assessment methods instructors apply to evaluate students in music classrooms. There is an ongoing trend of utilising peer and self-assessment methods and moving away from traditional exams for music evaluation (Hu et al., 2021). Previous studies have explored the benefits of collaborative learning and self-assessment in promoting student engagement and enhancing musical learning experiences (e.g., Ritchie & Williamon, 2012). This cluster is crucial for educators and policymakers as it emphasises adopting innovative and student-centred assessment approaches in music education to foster a supportive and inclusive learning environment.

Comprising nine keywords, cluster 3 includes terms like "Computer-aided instruction," "Elearning," and "Student engagement." The theme highlights the influence of recent technologies and innovations on music assessment methods. Researchers have observed a significant impact of technology on student evaluation practices in recent history (Sun, 2022). Previous studies have explored using e-learning platforms and computer-aided instruction to enhance student engagement and facilitate personalised learning experiences in music education (e.g., Vassallo & Rogers, 2019). Educators and policymakers should pay attention to this cluster as it presents opportunities for integrating technology and innovative tools into music assessment practices, fostering a technologically enriched learning environment.

Comprising eight keywords, the last cluster focuses on "Music education," "Teaching effect," "Vocal music," and "Education computing." The theme centres on the broader scope of music education and its relationship with assessment practices. Various teaching practices and evaluation methods have greatly influenced Chinese music education, especially by incorporating technology and other emerging trends (Liu et al., 2021). Previous studies have explored the impact of education computing and innovative teaching methods on music learning outcomes (e.g., Lee & Lim, 2019). This cluster holds significant relevance for educators, administrators, and policymakers, as it sheds light on the various factors influencing music education and highlights the need for continuous improvement in assessment strategies to cater to diverse student needs.

Music teachers strive to help students become independent musicians capable of critiquing their learning, work, and performance and improving based on feedback. Classroom assessment strategies that engage students in providing feedback to themselves and each other can create a shared responsibility for listening, critiquing, and revising and help students assume greater independence in and control over their learning (Liu & Careless, 2006). Educators have tested several formative classroom-based assessment strategies that include students as a critical source of feedback throughout the learning process (Chen et al., 2014).

The bibliometric analysis in the research was based on Chinese music education in higher education. The study found that assessment methods are changing according to the curricula and requirements of best evaluation techniques. Peer assessment and self-assessment are key terms that provide innovative solutions to music teachers in Chinese higher education (Yu & Leung, 2019). Through collaborative learning and student engagement, assessors expect to give a better evaluation to their students.

Self- and peer assessment were two processes of formative assessment that teachers were encouraged to use in their classrooms. In music, self-assessment is a critical element of effective independent practice (Cheng et al., 2020). During self-assessment, students critique their work according to explicitly stated expectations, usually in the form of goals or criteria, and then revise to improve their work. Self-assessment enhances the quality of first attempts at a piece of work so that the finished product or performance meets or exceeds expectations. Given this purpose, self-assessment is not self-evaluation, assigning a grade to one’s work. Rather, self-assessment is meant to allow students to take control over their learning by having them assess gaps in their understanding and skills and then use what they learn about their strengths and weaknesses as feedback for closing those gaps.

The process has been greatly influenced by technology and new educational concepts, and scholars believe that peer assessment and self-assessments are the best alternative methods to evaluate music students in higher education institutes in China (Li et al., 2021; Gong et al., 2022).

CONCLUSION

Research that sheds light on the power of formative assessment practices in music education is starting to emerge. In two recent studies of formative assessment in the arts (Andrade et al., 2014; Mastrorilli, 2014), music students had significantly higher music achievement when their teacher used

productive assessment practices similar to those described in this article. The bibliometric analysis was focused on Chinese music education in higher education. Findings concluded that peer and selfassessments are most effective for student engagement and provide a more practical and efficient approach to assessing the student.

The research sheds light on formative assessment practices in Chinese music education at the higher education level. While valuable insights were gained, limitations include the potential exclusion of relevant non-indexed research, a focus on quantitative co-occurrence analysis, and limited generalizability to other educational contexts. Nevertheless, the findings emphasise the effectiveness of peer and self-assessment methods in promoting student engagement and practical assessment approaches. Further research is warranted to deepen the understanding and implementation of these practices.

DATA AVAILABILITY DECLARATION

The original contributions encompassed within this study are comprehensively documented in the article and accompanying supplementary materials. Should additional inquiries or data-related requests arise, kindly direct them to the attention of the corresponding author.

FUNDING

This research endeavour did not receive financial support or grants from external sources.

ACKNOWLEDGEMENT

We sincerely thank our supervisor, [Supervisor Name], and colleagues, [Colleague Names], for their guidance and support. We also appreciate the assistance [Institution/Department] provided and the funding from [Funding Agency/Grant Number]. Special thanks go to our family and friends for their encouragement.

CONFLICT OF INTEREST

The authors herein assert that the research undertaken was executed without the influence of any commercial or financial affiliations, which may be perceived as potential conflicts of interest.

PUBLISHER'S NOTE

All assertions articulated in this article are exclusively attributable to the authors and may not necessarily reflect the views of their associated institutions, the publisher, the editors, or the reviewers. The evaluation of any product within this article or any claims made by its manufacturer does not entail a guarantee or endorsement by the publisher.

REFERENCES

Alessandri, E., Eiholzer, H., and Williamon, A. (2014). Reviewing critical practice: An analysis of Gramophone's reviews of Beethoven's piano sonatas, 1923-2010. Musicae Sci. 18, 131–149. doi: 10.1177/1029864913519466

Alessandri, E., Williamson, V. J., Eiholzer, H., and Williamon, A. (2015). Beethoven recordings reviewed: a systematic method for mapping the content of music performance criticism. Front. Psychol. 6:57. doi: 10.3389/fpsyg.2015.00057

Anglada-Tort, M., and Müllensiefen, D. (2017). The repeated recording illusion: the effects of extrinsic and individual difference factors on musical judgments. Music Percept. 35, 94–117. doi: 10.1525/mp.2017.35.1.94

Asch, S. E. (1956). Studies of independence and conformity. A minority of one against a unanimous majority. Psychol. Monogr. 70, 1–70. doi: 10.1037/h0093718

Aufegger and all. (2017). Musicians' perceptions and experiences of using simulation training to develop performance skills. Psychol. Music 45, 417–431. doi: 10.1177/0305735616666940

Bissonnette, J., Dubé, F., Provencher, M. D., and Sala, M. T. (2011). "The effect of virtual training on music performance anxiety," in Proceedings of the International Symposium on Performance Science 2011, eds A. Williamon, D. Edwards, and L. Bartel (Utrecht: European Association of Conservatoires), 585–590.

Blascovich, and al (2002a). Immersive Virtual Environment technology as a methodological tool for social psychology. Psychol. Inq. 13, 103–124. doi: 10.1207/S15327965PLI1302_01

Blascovich, and all (2002b). Immersive Virtual Environment technology: just another methodological tool for social psychology? Psychol. Inq. 13, 146–149. doi: 10.1207/S15327965PLI1302_03

Bliss, and all (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. Presence 6, 73–86. doi: 10.1162/pres.1997.6.1.73

Blom, D., and Encarnacao, J. (2012). Student-chosen criteria for peer assessment of tertiary rock groups in rehearsal and performance: What is essential? Br. J. Music Educ. 29, 25–43. doi: 10.1017/S0265051711000362

Blom, D., and Poole, K. (2004). Peer assessment of tertiary music performance: Opportunities for understanding performance assessment and performing through experience and self-reflection. Br. J. Music Ed. 21, 111–125. doi: 10.1017/S0265051703005539

Bonneville-Roussy, A., Bouffard, T., and Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. J. Sch. Psychol. 63, 1–12. Doi: 10.1016/j.jsp.2017.02.002

Bonshor, M. (2017). Conductor feedback and the amateur singer: the role of criticism and praise in building choral confidence. Res. Stud. Music Educ. 39, 139–160. doi: 10.1177/1321103X17709630

Bouchard, and all (2006). Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. Technol. Health Care 14, 19–27.

Chanwimalueang, and all (2017). Stage call: cardiovascular reactivity to audition stress in musicians. PLoS ONE 12:e0176023. Doi: 10.1371/journal.pone.0176023

Daniel, R. (2001). Self-assessment in performance. Br. J. Music Educ. 18, 215–226. doi: 10.1017/S0265051701000316

Daniel, R. (2004). Peer assessment in musical performance: The development, trial and evaluation of a methodology for the Australian tertiary environment. Br. J. Music Educ. 21, 89–110. doi: 10.1017/S0265051703005515

de Borst, and all (2015). Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. Front. Psychol. 6:576. doi: 10.3389/fpsyg.2015.00576

Difede, J., Cukor, J., Jayasinghe, N., Patt, I., Jedel, S., Spielman, L., et al. (2002). Virtual reality exposure therapy for World Trade Center post-traumatic stress disorder: a case report. Cyberpsychol. Behav. 5, 529-535. doi: 10.1089/109493102321018169

Dotger, and all (2018). In the limelight: utilising clinical simulations to enhance music teacher education. Res. Stud. Music Educ. doi: 10.1177/1321103X18773102. [Epub ahead of print].

Duerksen, G. L. (1972). Some effects of expectation on evaluation of recorded musical performance. J. Res. Music Educ. 20, 268–272. doi: 10.2307/3344093

Duke, R. A. (1999). Measures of instructional effectiveness in music research. Bull. Council Res. Music Educ. 143, 1–48.

Eggleston, J. (1991). Teaching teachers to assess. Eur. J. Educ. 26, 231–237. doi: 10.2307/1503025

Elliott, C. A. (1995). Race and gender as factors in judgments of musical performance. Bull. Council Res. Music Educ. 127, 50–56.

Endo, S., and all (2014). Interaction between physiological and subjective states predicts the effect of a judging panel on the postures of cellists in performance. Front. Psychol. 5:773. doi: 10.3389/fpsyg.2014.00773

Ericsson, and al (1993). The role of deliberate practice in the acquisition of expert performance. Psychol. Rev. 100, 363–406. doi: 10.1037/0033-295X.100.3.363

Falchikov, and all (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. Rev. Educ. Res. 70, 287–322. doi: 10.3102/00346543070003287

Ferm Almqvist, C., Vinge, J., Väkevä, L., and Zandén, O. (2016). Assessment learning in music education: the risk of "criteria compliance" replacing "learning" in the Scandinavian countries. Res. Stud. Music Educ. 39, 3–18. doi: 10.1177/1321103X16676649

Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. J. Res. Music Educ. 23, 186–196. doi: 10.2307/3344643

Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. J. Res. Music Educ. 25, 256–263. doi: 10.2307/3345266

Gaunt, H. (2017). "Apprenticeship and empowerment: the role of one-to-one lessons," in Musicians in the Making: Pathways to Creative Performance, eds J. Rink, H. Gaunt, and A. Williamon (Oxford: Oxford University Press), 28–56.

Glowinski et al (2015). Evaluating music performance and context-sensitivity with Immersive Virtual Environments. EAI Endorsed Trans. Creat. Technol. 2:e3. doi: 10.4108/ct.2.2e3

Goolsby, T. W. (1999). Assessment in instrumental music. Music Educ. J. 86, 31–50. doi: 10.2307/3399587

Griffiths, N. K. (2008). The effects of concert dress and physical appearance on perceptions of female solo performers. Musicae Sci. 12, 273–290. doi: 10.1177/102986490801200205

Griffiths, N. K. (2010). 'Posh music should equal posh dress': an investigation into female soloists' concert dress and physical appearance. Psychol. Music 38, 159–177. doi: 10.1177/0305735608100372

Hewitt, M. P. (2015). Self-efficacy, self-evaluation, and music performance of secondary-level band students. J. Res. Music Educ. 63, 298–313. doi: 10.1177/0022429415595611

Hunter, D., and Russ, M. (1996). Peer assessment in performance studies. Br. J. Music Educ. 13, 67. doi: 10.1017/S0265051700002953

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., and Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med. Teach. 27, 10–28. doi: 10.1080/01421590500046924

Johnston, H. (1993). Video self-assessment, peer assessment, and instructor feedback are used to evaluate conducting skills in music student teachers. Br. J. Music Educ. 10, 57. doi: 10.1017/S0265051700001431

Jørgensen, H. (2004). "Strategies for individual practice," in Musical Excellence: Strategies and Techniques to Enhance Performance, ed A. Williamon (Oxford: Oxford University Press), 85–104.

Jørgensen, H. (2008). Instrumental practice: quality and quantity. Finnish J. Music Educ. 11, 8–18.

Kassab, and all (2011). "Blowing up the barriers" in surgical training: exploring and validating the concept of distributed simulation. Ann. Surg. 254, 1059–1065. doi: 10.1097/SLA.0b013e318228944a

Kätsyri, and all. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Front. Psychol. 6:390. doi: 10.3389/fpsyg.2015.00390

Kenny, D. T. (2011). The Psychology of Music Performance Anxiety. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199586141.001.0001

Kneebone and all. (2010). Distributed simulation–accessible, immersive training. Med. Teach, 32, 65–70. doi: 10.3109/01421590903419749

Kolb, A. Y., and Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in Higher Education. Acad. Manag. Learn. Educ. 4, 193–212. doi: 10.5465/amle.2005.17268566

Kopiez, and all (2017). Small influence of performing from memory on audience evaluation. Empir. Musical. Rev. 12, 2–14. doi: 10.18061/emr.v12i1-2.5553

Latukefu, L. (2010). Peer assessment in tertiary level singing: changing and shaping culture through social interaction. Res. Stud. Music Educ. 32, 61–73. doi: 10.1177/1321103X10370091

Lebler, D. (2007). Student-as-master? Reflections on a learning innovation in popular music pedagogy. Int. J. Music Educ. 25, 205–221. doi: 10.1177/0255761407083575

Lerman, L., and Borstel, J. (2003). Critical Response Process: A Method for Getting Useful Feedback on Anything You Make, From Dance to Dessert. Liz Lerman Dance Exchange.

Long, M., Creech, A., Gaunt, H., Hallam, S., and Robertson, L. (2012). Blast from the past: conservatoire students' experiences and perceptions of public master classes. Music. Sci. 16, 286–306. doi: 10.1177/1029864912458848

McCormick, J., and McPherson, G. (2003). The role of self-efficacy in a musical performance examination: An exploratory structural equation analysis. Psychol. Music 31, 37–51. doi: 10.1177/0305735603031001322

McPherson, G. E., and McCormick, J. (2006). Self-efficacy and music performance. Psychol. Music 34, 322–336. doi: 10.1177/0305735606064841

McPherson, G. E., and Schubert, E. (2004). "Measuring performance enhancement in music," in Musical Excellence: Strategies and Techniques to Enhance Performance, ed A. Williamon (Oxford: Oxford University Press), 61–82.

Medland, E. (2015). Examining the assessment literacy of external examiners. Lond. Rev. Educ. 13, 21–33. doi: 10.18546/LRE.13.3.04

Miles, H. C., Pop, S. R., Watt, S. J., Lawrence, G. P., and John, N. W. (2012). A review of virtual environments for training in ball sports. Comput. Graph. 36, 714–726. doi: 10.1016/j.cag.2012.04.007

Mitchell, H. F., and Benedict, R. (2017). The moot audition: preparing music performers as expert listeners. Res. Stud. Music Educ. 39, 195–208. doi: 10.1177/1321103X17709631

Nicol, D. J., and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Stud. Higher Educ. 31, 199–218. doi: 10.1080/03075070600572090

Nieuwenhuys, A., and Oudejans, R. (2012). Anxiety and perceptual-motor performance: Toward an integrated model of concepts, mechanisms, and processes. Psychol. Res. 76, 747–759. doi: 10.1007/s00426-011-0384-x

Orman, E. K. (2003). Effect of virtual reality graded exposure on heart rate and self-reported anxiety levels of performing saxophonists. J. Res. Music Educ. 51, 302–315. doi: 10.2307/3345657

Orman, E. K. (2004). Effect of virtual reality graded exposure on anxiety levels of performing musicians: a case study. J. Music Ther. 41, 70–78. doi: 10.1093/jmt/41.1.70

Paris, S. G., and Winograd, P. (1990). "How metacognition can promote academic learning and instruction," in Dimensions of Thinking and Cognitive Instruction, eds B. F. Jones and L. Idol (Hillsdale, NJ: Lawurence Erlbaum Associates, Inc.), 15–51.

Perkins, R. (2013). Learning cultures and the conservatoire: An ethnographically-informed case study. Music Educ. Res. 15, 196–213. doi: 10.1080/14613808.2012.759551

Platz, F., and Kopiez, R. (2012). When the eye listens: a meta-analysis of how audio-visual presentation enhances the appreciation of music performance. Music Percept. 30, 71–83. doi: 10.1525/mp.2012.30.1.71

Platz, F., and Kopiez, R. (2013). When the first impression counts: Music performers, audience and the evaluation of stage entrance behaviour. Musicae Sci. 17, 167–197. doi: 10.1177/1029864913486369

Radocy, R. E. (1976). Effects of authority figure biases on changing judgments of musical events. J. Res. Music Educ. 24, 119–128. doi: 10.2307/3345155

Ravel, M. (1919). Le tombeau de Couperin: Suite d'orchestre (Orchestral Score: 1st & 2nd Oboe). Paris: Durands and Fils.

Ritchie, L., and Williamon, A. (2011). Measuring distinct types of musical self-efficacy. Psychol. Music 39, 328–344. doi: 10.1177/0305735610374895

Ritchie, L., and Williamon, A. (2012). Self-efficacy as a predictor of musical performance quality. Psychol. Aesthet. Creat. Arts 6, 334–340. doi: 10.1037/a0029619

Robinson, C. R. (1993). Singers' self-assessment of choral performance: next-day recollections versus concert tape evaluation. Southeastern J. Music Educ. 4, 224–233.

Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., and Price, L. (2000). A controlled study of virtual reality exposure therapy for fear of flying. J. Consult. Clin. Psychol. 68, 1020–1031. doi: 10.1037/0022-006X.68.6.1020

Russell, B. E. (2015). An empirical study of a solo performance assessment model. Int. J. Music Educ. 33, 359–371. doi: 10.1177/0255761415581282

Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. Nat. Rev. 6, 332–339. doi: 10.1038/nrn1651

PubMed Abstract | CrossRef Full Text | Google Scholar

Searby, M., and Ewers, T. (1997). An evaluation of peer assessment in higher education: a case study in the School of Music, Kingston University. Assess. Eval. Higher Education. 22, 371–383. doi: 10.1080/0260293970220402

Slater, M., Pertaub, D.-P., and Steed, A. (1999). Public speaking in virtual reality: facing an audience of avatars. IEEE Comput. Graph. Appl. 19, 6–9. doi: 10.1109/38.749116

Stewart, A. (2011). Examiner Training: The Full Story. Available online at: https://us.abrsm.org/en/exam-support/exam-support-articles/article/examiner-training-thefull-story/175/ (accessed June 5, 2018).

Szigeti, J. (1947). With Strings Attached: Reminiscences and Reflections. New York, NY: Alfred A. Knopf.

Taylor, A. (2010). Participation in a master class: experiences of older amateur pianists. Music Educ. Res. 12, 199–217. doi: 10.1080/14613801003746576

Tchaikovsky, P. I. (1946). Symphony No. 4 in F Minor, Op. 36 (Orchestral Score: 1st Oboe). Leipzig: Bruckner-Verlag.

Thompson, S., and Williamon, A. (2003). Evaluating evaluation: Musical performance assessment as a research tool. Music Percept. 21, 21–41. doi: 10.1525/mp.2003.21.1.21

Thompson, S., Williamon, A., and Valentine, E. (2007). Time-dependent characteristics of performance evaluation. Music Percept. 25, 13–29. doi: 10.1525/mp.2007.25.1.13

Waddell, and all (2018). Making an impression: Error location and repertoire features affect performance quality rating processes. Music Percept. 36, 60–76. doi: 10.1525/mp.2018.36.1.60 Waddell et al (2017a). “Measuring the audience,” in Scholarly Research for Musicians, ed S. H. Lee (NY: Routledge), 148–155.

Waddell, G., and Williamon, A. (2017b). Eye of the beholder: Stage entrance behavior and facial expression affect continuous quality ratings in music performance. Front. Psychol. 8:513. doi: 10.3389/fpsyg.2017.00513

Wesolowski, and all (2015). Rater fairness in music performance assessment: evaluating model-data fit and differential rater functioning. Musicae Sci. 19, 147–170. doi: 10.1177/1029864915589014

Wesolowski, and all (2016). Examining rater precision in music performance assessment: an analysis of rating scale structure using the multifaceted Rasch Partial Credit Model. Music Percept. 33, 662–678. doi: 10.1525/mp.2016.33.5.662

Williamon, and all (2014). Simulating and stimulating performance: introducing distributed simulation to enhance musical learning and performance. Front. Psychol. 5:25. doi: 10.3389/fpsyg.2014.00025

Winter, N. (1993). Music performance assessment: a study of the effects of training and experience on the criteria used by music examiners. Int. J. Music Educ. 22, 34–39. doi: 10.1177/025576149302200106

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: an overview. Educ. Psychol. 25, 3–17. doi: 10.1207/s15326985ep2501_2.