

MUSIC PERFORMANCE EVALUATION IN TEACHERS' COLLEGES IN JAMAICA

O'Neal Mundle

There is need to maintain quality in the offerings of courses at teachers' colleges in Jamaica, especially in the area of music performance, for which appreciation is largely subjective and low in inter-rater reliability. The purpose of this study was to determine the performance of students in practical music examinations at teachers' colleges in Jamaica. Practical music examination grades of 226 students in the Primary and Early Childhood diploma programmes of five teachers' colleges were analysed. Comparisons were made between grades obtained in external and internal practical examinations; similarly, comparisons were made between grades assigned by Early Childhood programmes and those earned in Primary programmes. No statistically significant difference was found between the grades assigned by internal examiners and those assigned by external examiners. Overall, most students were successful in the course, and more grades assigned by internal examiners were maintained during external assessments than were downgraded or upgraded, with downgrades mainly occurring for students in Early Childhood teacher training programmes. Disparities between external examiners' grades and internal examiners' grades were largely limited to a few colleges. Finally, the study presents perspectives of internal practical examiners on the quality assurance procedure used in the practical music examinations.

Introduction

The need for quality assurance in maintaining standards is particularly relevant to disciplines such as music, for which appreciation is subjective by nature. In fact, the need for independent assessments of performances is especially relevant to musical performances. Gentles (2011) explained that for many years external examiners had been integral to this quality assurance process, as carried out through the Joint Board of Teacher Education (JBTE) since its inception in the 1960s.

Quality Assurance

Ross (2009) traced the use of external examiners to 1832 at the University of Durham, and argued that external music examiners act in the capacity of assessors, evaluators, or advisors in the process of maintaining musical standards. They are an important means of quality assurance, and represent “a measure of comparable standards from an outside perspective” (p. 474). Some are appointed by an examining board on an ongoing or ad hoc basis. Many British centres engage external examiners. For example, in 2002, the Associated Board of the Royal Schools of Music used a team of 750 external music examiners, who were responsible for maintaining quality with 100,000 teachers, who in turn taught 600,000 candidates. Ross further argued that as quality assurance practices have become more heavily scrutinized, often influencing institutional rankings, external music examiners provide valuable feedback to course instructors and accreditation boards, encouraging a culture of transparency in music assessment (p. 474).

Reliability in Music Performance Evaluation

Given the subjective nature of appreciation for the arts, there has been some difficulty in achieving reliability in the evaluation of performances in music. Accuracy in music evaluation has been described as elusive (Bergee, 1997; Kinney, 2009). In fact, Bergee (1997) found a general lack of reliability among some faculty evaluators of music performances, and questions were consequently raised regarding the viability of such assessment. In contrast, later research by Bergee (2003) found inter-judge reliability to be stable and consistent among music evaluators.

There are extra-musical influences on the assessment process, and the validity of performance evaluation can sometimes be called into question. For example, attractiveness bias was found in favour of more attractive female pianists who were among the best, and also in favour of less attractive male pianists (Ryan & Costa-Giomi, 2004). Also, a bias towards whites was found in a study where race was considered (Elliot, 1995/1996). Thompson and Williams (2003) recognized implicit biases by adjudicators who admitted difficulty in assessing performances on instruments other than their own. Audition scores for an all-state band were better for students from larger schools and for those who travelled a shorter distance to the audition site (Lien & Humphreys, 2001). Other important non-musical influences on performance evaluation include: time of day; size of school; performing medium (vocal or instrumental); and type of event (solo or

ensemble); with ratings, for example, moving towards *superior* as the day progressed (Bergee & McWhirter, 2005; Bergee & Platt, 2003).

This study therefore seeks to build on existing research that has determined reliability in the evaluation of music performances, and is situated within the context of the need for reliability between internal and external examiners' evaluation of musical performances by students in the teachers' colleges of Jamaica. The purpose of the study is to assess the performance of teachers' college students in the practical component of their music examinations, as well as to determine the level of variability between the evaluations of internal and external examiners. The study also sought to determine whether there were relationships in the performances based on sex, level of the candidates, or college attended. Finally, I explored the perspectives of college examiners on the quality assurance process as used in the external music practical examinations.

Research Questions

- 1. What are the grades attained by students in practical music examinations at selected teachers' colleges between 2011 and 2013?*
- 2. Is there a significant difference in grades assigned in internal exams and grades assigned by external examiners?*
- 3. What is the relationship between sex and the adjustments to grades?*
- 4. What is the relationship between the college attended and grade adjustments?*
- 5. Is there a relationship between the level of the programme (Early Childhood and Primary programmes) and the grades assigned students?*
- 6. What are the perspectives of internal examiners on the quality assurance process used in the music practical examinations?*

Research Context

For many years, Jamaica has had less than 10 institutions responsible for the training of general classroom teachers at the primary and early childhood levels. These institutions are: Sam Sharpe Teachers' College, Bethlehem Teachers' College, Moneague Teachers' College, Church Teachers' College, Shortwood Teachers' College, St. Joseph's Teachers' College, College of Agriculture and Science Education (CASE), and Mico University College. The Edna Manley

College of the Visual and Performing Arts and the G.C. Foster College of Physical Education and Sport focus on specialist areas: the arts in the case of the former and sport in the latter. Other more recent institutions offering pre-secondary level classroom training include the International University of the Caribbean and the Catholic College of Mandeville. Virtue (2013) reports that enrolment in teachers' colleges is largely female, with female students outnumbering males by ratios as high as 16:1. In 2012 at Shortwood Teachers' College, there were 838 women and only 57 men; and at Sam Sharpe Teachers' College there were 734 females and 122 males in training.

Tucker (1995) explains that with the integrated curriculum introduced by the Jamaican Government in the 1970s, pre-secondary general classroom teachers began acquiring, through teachers' colleges, the skills and knowledge required to impart music at that level. As part of their course of study in the present scenario, pre-service teachers are required to complete a one-semester general music course intended to expose them to basic classroom music. Quality assurance for the course is governed through the JBTE, which has among its responsibilities, the maintenance and enhancement of academic standards. According to the JBTE (2013):

A system of quality assurance acts as a catalyst for internal improvements within a context of advice, support and dialogue in an environment of trust. Among the variety of strategies used is a large system of external examiners which provides one of the principal means for maintaining nationally comparable standards among education institutions offering programmes at a similar level. The external examiners oversee the preparation of syllabuses, the marking of examinations and the assessment of the practicum exercise. (Quality Assurance Responsibility)

Summative assessment for the General Music Course takes the form of a written (theory-based) component and a practical component. The areas examined within the practical examination are: melodic and rhythmic sight-reading, aural and rhythmic recall, the performance of prepared pieces on the recorder, and the playing of scales in major and minor modes on the recorder (Joint Board of Teacher Education, 2004).

Quality Assurance Assessment Procedures

The quality assurance process involves an internal assessment examination conducted by college music lecturers. Within a short time, usually no more than a week, students are required to perform

the same examination for external examiners in the external examination. The external examination team usually consists of the Chief External Examiner and other music examiners who are practitioners in the field of music. For both internal and external examination exercises, candidates perform two prepared pieces from their repertoire on the recorder worth 40%, in addition to presenting selected scales in the major and minor modes (20%). Candidates also play at sight, short, rhythmic, and melodic excerpts not previously seen (24%); and are required to recall and perform short rhythmic and melodic patterns as presented by the examiner (16%).

Method

Measures

Examiners assign evaluation grades to each performance, which range through: A, B+, B, B-, C+, C, C-, D, and F, where an A represents the highest grade and D and F represent a failing grade. For the purpose of this study, comparisons between the internal and external grades are categorized as: *Upgraded* – when the external examiner raises the grade; *Downgraded* – when the grade is lowered in the external examination; and *Maintained* – when the same grade is maintained. The names, identification numbers, or any other identifying feature of individual candidates were not included in the data, and permission was obtained from the JBTE to use these data in the present study.

Sample

The population/sample consisted of 226 students in the diploma programmes from selected teachers' colleges: St. Joseph's, Moneague, Church, Bethlehem, and CASE. These students were examined over a two-year period. There were 213 female and 13 male subjects, all pursuing the primary and early childhood teacher education programmes. The disproportionately low number of male subjects when compared to female subjects is reflective of enrolment patterns in primary and early childhood education at the teachers' colleges in Jamaica.

As non-music specialists, they were required to engage in a music course designed to expose teachers to general classroom music. The course required that subjects do a practical music examination. Close to the end of the two-year quantitative data collection period, interviews were conducted with each of the five college lecturers/internal examiners associated with the respective colleges. The interviews sought to ascertain experiences with the quality assurance procedure,

strengths and weaknesses, and suggestions for improving the procedure. The following questions guided the interviews:

1. How important are the external music practical examinations?
2. What were the experiences of your students in the external music practical examinations?
3. What do you consider as areas of weaknesses in the external music practical examination process?
4. How can the external music practical examinations be improved?
5. How can the quality assurance process for music external examinations be improved?

Analytical Procedures

The quantitative dimension of the research utilized *t*-tests and chi-square analyses. Letter grades were converted to weighted numbers to facilitate quantitative analysis. The grades assigned were converted to a 9-point scale (F=1, D=2 etc., to A=9). For the qualitative analysis, the researcher read and analysed the interview transcripts, assigning themes to paragraphs, phrases, and sentences to reflect the meanings of the interview responses.

Findings

Internal vs. External Scores

An independent samples *t*-test was conducted to compare the students' music scores, as seen in Table 1. With equality of variance assumed, there was no statistical significant difference between the internal (M=6.15, SD=1.95) and external scores (M=6.26, SD =2.198); $t(225) = -1.065$, $p = .288$ (two tailed).

Table 1. Independent t-test Results Comparing Scores by the Internal and External Examiners' Grades

	Modes of examination		<i>t</i>	<i>df</i>
	Internal	External		
Music Scores	6.15	6.26	-1.065	225
	(1.95)	(2.19)		

Note. * = $p \leq .05$, ***= $p \leq .001$. Standard deviations appear in parentheses below means.

Figure 1 shows that the highest proportion of grades were maintained (41%), while 23% of grades were downgraded, and 37% were upgraded.

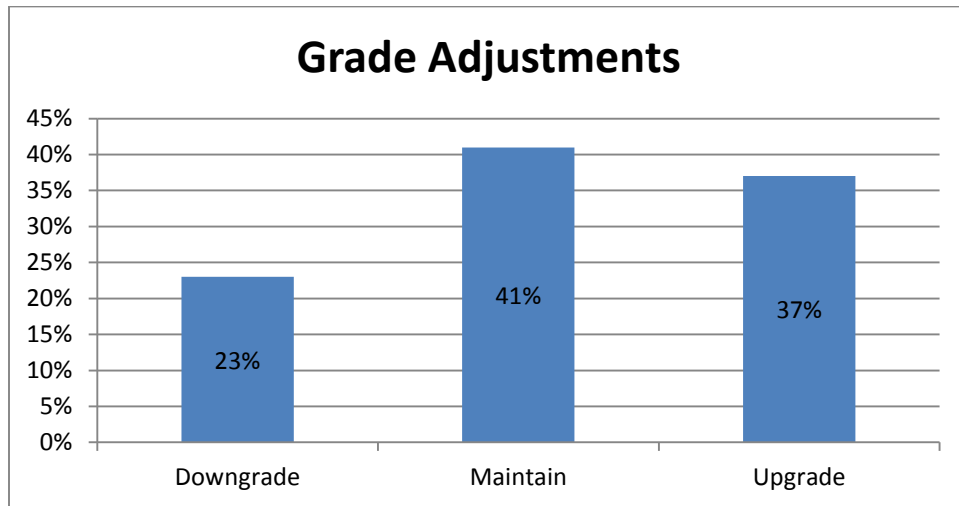


Figure 1. Bar graph showing direction of grade adjustments in percentages.

A crosstabulation chi-square was used to examine the association between sex and grade adjustment, as seen in Figure 2. The results showed a positive but weak relationship between both variables; $\chi^2 (2) = 6.37$, $p (.041) < 0.05$; $cc .16$. The general pattern is that males are two times more likely to experience an upgrade in their music scores when their grades are adjusted. This pattern must, however, be viewed in the context that the sample of males is relatively small (13 males), posing a difficulty in drawing firm conclusions on their performance versus that of females.

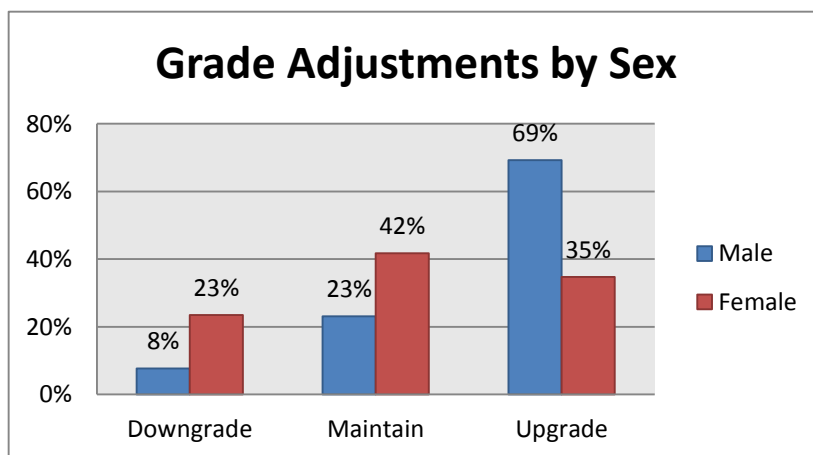


Figure 2. A crosstabulation chi-square examining the association between gender and grade adjustment.

Crosstabulation chi-square was used to examine the association between music level and grade adjustment. As seen in Figure 3, the results showed a positive but weak relationship between both variables; $\chi^2(2) = 34.45, p(.000) < 0.01; cc .364$. The general pattern is that the internal scores for persons at the primary level are more likely to be maintained.

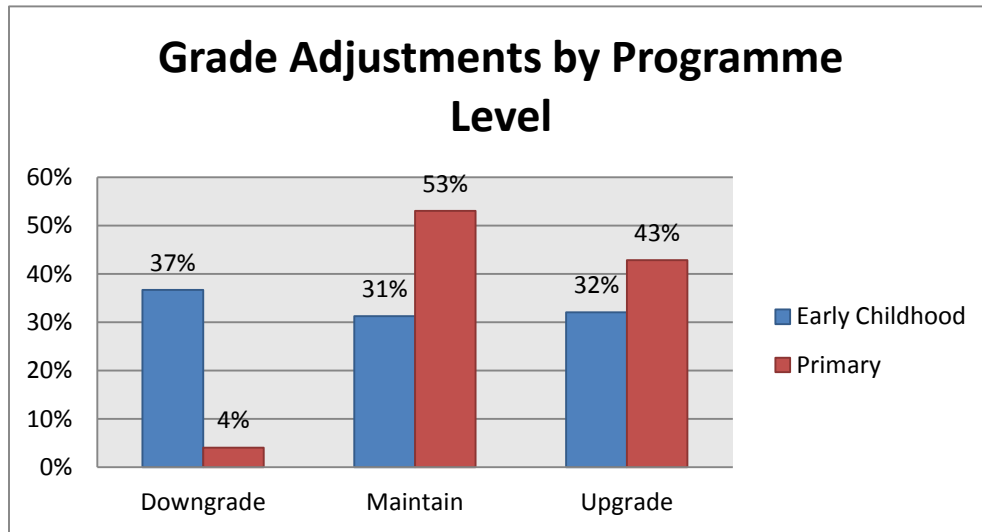


Figure 3. Bar chart showing the association between music level and grade adjustment.

Figure 4 shows the relationship between the adjustments to internal grades and the college that the candidate attends.

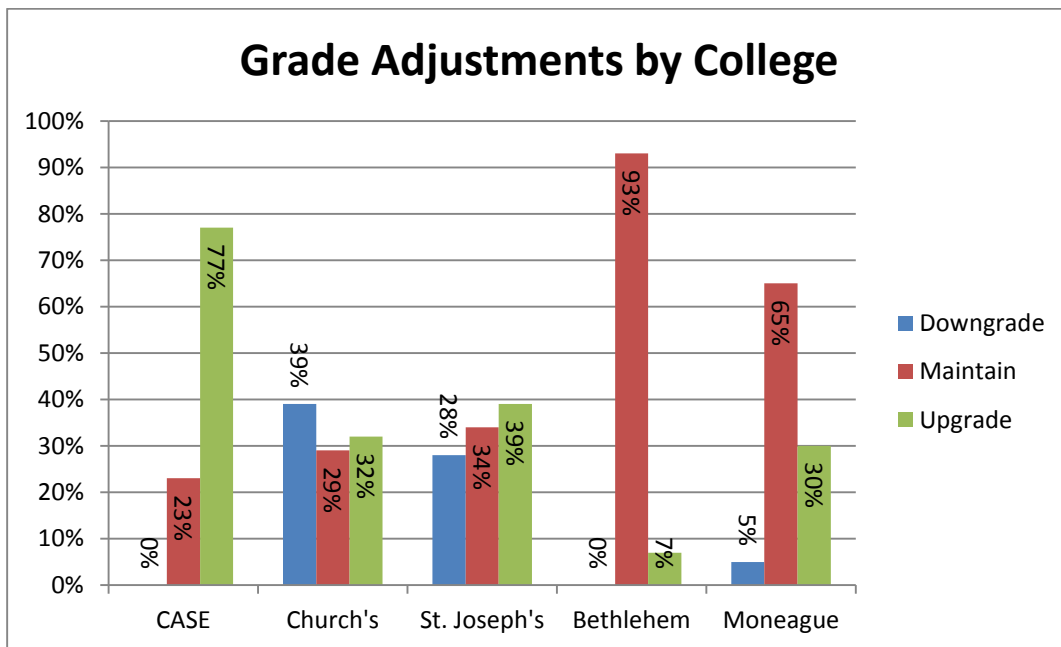


Figure 4. Bar graph showing the association between college attended and grade adjustment.

A crosstabulation chi-square was used to examine the association between college attended and grade adjustment. As seen in Figure 4, the results showed a positive moderate relationship between both variables; $\chi^2 (2) = 68.56$, $p (.000) < 0.01$; $cc .48$. The general pattern is that the internal score for candidates at the Bethlehem Teachers' College is most likely to be maintained, while CASE obtained the most upgrades, and Church Teachers' College had the highest number of downgrades.

Figures 5 and 6 show how grades are distributed by the internal examiners at the colleges, and how grades are distributed when evaluated by the external examiners. For internal exams, just under half of the students (49%) were evaluated as earning a grade of B and above, while external examiners awarded grades B and above to a little more than half of the candidates (53%).

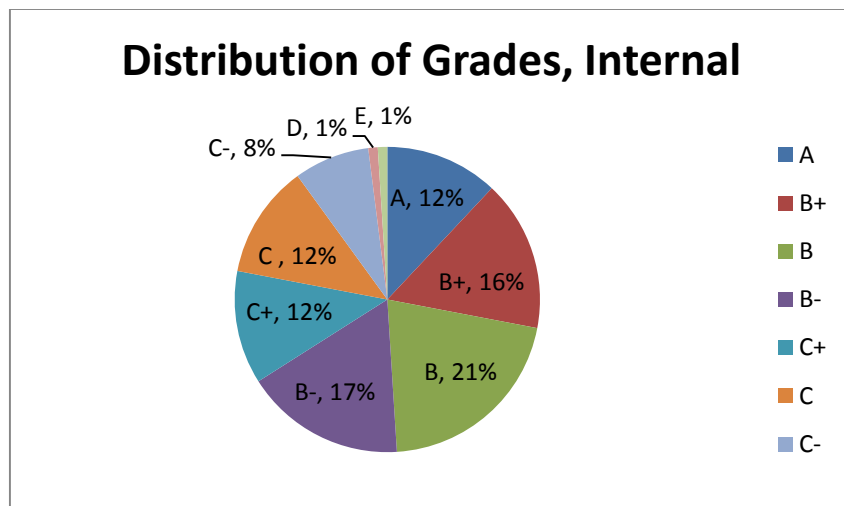


Figure 5. Pie chart showing how the internal examiners distributed the grades.

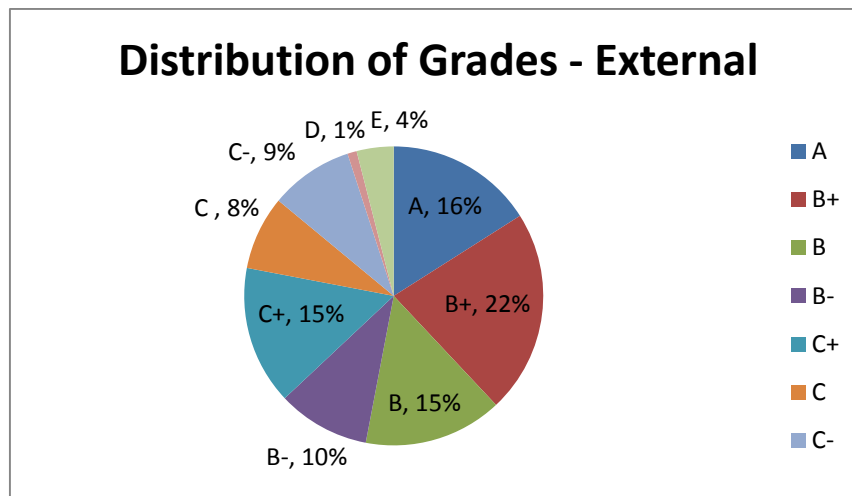


Figure 6. Pie chart showing how the external examiners distributed the grades.

External music practical examinations are important for setting and maintaining standards. The lecturers generally agreed that the process was useful as a quality assurance tool. In the words of one instructor: *“It helps lecturers to maintain best practices, and it helps to assure parity and fairness across the standards at participating institutions.”* Another lecturer said: *“When it is the same examination team that does it in other places, I am confident that my grades are on par with the other colleges.”*

Improvement in performance between internal and external examinations. There was a common view that students performed better on the external examinations than they did on the internal. This improvement was credited to the external process being perceived as more important to the students. Also, instructors pointed out that they tend to give improvement tips subsequent to the internal examination and prior to the external examinations. Lecturers also opined that they probably mark the students too hard, therefore the input of the external examiners in determining the overall practical grade was a clear positive.

Student reactions. Some lecturers reported that their students would get very nervous and tense at the thought of participating in an external music practical examination. They were, however, usually excited after the external exams. Says a lecturer: *“The students usually report that the external examiners are not really monsters. I feel that as their lecturers, we are harder on them, than the externals.”*

Gap between time of internal and external exams. Lecturers had concerns about the ensuing gap between the occurrence of the internal and external practical evaluations. They argued that it should be consistent for all colleges. Says one lecturer: *“An established examination period should be given and strictly adhered to. For example, one week before or after written examinations, and one week after the internal.”* Another suggestion is that arrangements should be made for the external and internal evaluations to be done simultaneously and moderated at the same time.

Need for greater standardization. It was felt that the rubric and scoring instruments may need to be more specific, and guidelines on procedures, including consequences, should be given and agreed on by external and internal examiners to ensure fairness and understanding of the process. Even though there are guidelines for the practical examinations outlined in the syllabus,

there is need for standardization exercises to be done. For example, selected pieces for each examination should be agreed upon and circulated. Argues a lecturer: *“To my knowledge, there has never been an exercise of this nature carried out.”*

Feedback on evaluations. It was suggested that detailed collective and individual reporting on the process and outcomes by the external team is needed after each exercise. Reports should include strengths, weaknesses, suggestions, and recommendations. One lecturer pointed out that *“in the past, reports have been submitted by External Examiners but they are not always shared with college lecturers.”*

Quality assurance through lecturer exchange. Two lecturers explored the idea of an exchange of examiners across colleges, where lecturers would visit each other's colleges to carry out the quality assurance process.

Discussion of Findings

The results of the internal and external music performance evaluations/examinations revealed that, over the two-year period, the large majority of students who were evaluated successfully executed the music performance component of the course. For both internal and external examinations, approximately half of the students obtained a grade in the B range, suggesting that most candidates gave an average performance; and relatively few candidates were evaluated as giving *extremely good* or *extremely weak* performances. Interestingly, the external examiners were more likely to award a grade of A (16%) than the internal examiners (12%).

This study shows that males are twice as likely to experience an upgrade in their music scores when their grades are adjusted in the external examination. It is of significance, however, that of the 226 pre-service teachers evaluated, only 13 were male. This scenario is consistent with trends in the demographic breakdown of teachers in the Jamaica, where relatively few of the teachers are male, and even less so at the pre-secondary levels. In most teacher training institutions in Jamaica, females outnumber males by as much as 16:1, except at the Edna Manley College of the Visual and Performing Arts, where the gap is much closer; and the G.C. Foster College of Physical Education and Sport, where there are more males than females (Virtue, 2013).

Results of this study revealed no statistical significant difference between the internal and external scores, and, overall, the highest proportion of grades was maintained (41%). Such a

finding is consistent with research by Bergee (2003), where evaluation of musical performances indicated stable and consistent inter-rater reliability. The findings of this study are however in contrast to an earlier study by Bergee (1997), where reliability across evaluators was very low and elusive.

It is noteworthy that many of the grades (37%) were upgraded when the external examinations were conducted. Explanations for this phenomenon included the fact that the student performances could have genuinely improved between the time of the sitting of the internal and external exams due to the Hawthorne effect or a general familiarity with the examination process. Another explanation could be that the candidates made use of the practice time between the internal and external examinations, as was confirmed in the interviews with their instructors. Notwithstanding, it is well appreciated that musical skills are not acquired or lost overnight, so reliable evaluations can still be made regarding the musicianship of a person even though performances take place at two different times.

With regard to adjustments made to scores at the actual colleges, results showed greatest reliability at Bethlehem and Moneague Teachers' Colleges where 93% of internal scores were maintained for the former, and 65% maintained for the latter. Candidates at CASE obtained the most upgrades (78%), while those at Church Teachers' College had the highest incidence of downgrades (40%) compared with a similar level of score upgrades for St Joseph's Teachers' College. It is clear that a few colleges account for most of the changes in grades and, from the perspective of a quality assurance objective, there may be greater concern for cases where the grades appear inflated and should be downgraded, than for the reverse. The process, however, needs to guard against internal examiners being deliberately too conservative when carrying out evaluations for fear that the quality assurance process may reflect negatively on them if too many of their grades are adjusted downward. In the interviews, internal examiners in fact did voice the possibility that they were too "*hard on their students.*"

Findings from this study include some valuable recommendations, particularly the need to ensure standardization of the procedures, especially as it relates to the work of the internal examiners. Such standardization ranges from the choosing of pieces of comparable difficulty levels across colleges, to having the same time lapse between internal and external examinations. Another recommendation that should improve the process, if it is feasible, is that of merging the internal and external exams to the same day. From the point of view of reliability in evaluation, such a

move would remove external variability to what is presented to the examiners for adjudication. It may not, however, result in the students being better off for the experience, as they would only be afforded one playing opportunity, with no second chance.

Further Research

Further research should focus on comparisons between the performance of students in the overall music examinations (practical and theory components) and in other subjects that they engage in during teacher preparation. Another area of enquiry should seek to document the experiences of students in the music examinations, both internal and external. Finally, future research should also seek the views of external examiners on how the quality assurance process could be most efficiently implemented to be responsive to the peculiarities of evaluating music performances.

Conclusion

In conclusion, pre-service primary and early childhood teachers do reasonably well in practical music exams that they are required to sit, with most obtaining passing grades. There are also high levels of agreement between the evaluations of the internal and external examiners. It appears that students are more likely than not to benefit from the review process that the external examination affords, as most adjustments to grades were in an upward direction. The quality assurance process is therefore valuable, and its integrity not in question, notwithstanding areas that could be improved.

References

- Bergee, M. J. (1997). Relationships among faculty, peer, and self-evaluations of applied performances. *Journal of Research in Music Education*, 45(4), 601–612. doi:10.2307/3345425
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51(2), 137–150. doi:10.2307/3345847
- Bergee, M. J., & McWhirter J. L. (2005). Selected influences on solo and small-ensemble festival ratings: Replication and extension. *Journal of Research in Music Education*, 53(2), 177–190. doi:10.2307/3345517
- Bergee, M. J., & Platt, M. C. (2003). Influence of selected variables on solo and small-ensemble festival ratings. *Journal of Research in Music Education*, 51(4), 342–353. doi:10.2307/3345660

- Elliot, C. A. (1995/1996). Race and gender as factors in judgments of musical performance. *Bulletin of the Council for Research in Music Education*, 127, 50–56.
- Hordatt Gentles, C. (2011). The external examiner in the context of a strengthened quality assurance environment. In M. A. Stewart (Ed.), *Quality assurance in Caribbean teacher institutions: Issues and perspectives* (Institute of Education Publication Series, 7; pp. 21–34). Mona, Jamaica: Institute of Education, UWI.
- Joint Board of Teacher Education. (2004). *Curriculum guide for three year Diploma in Teaching: Music – Early Childhood*. Retrieved from <http://www.jbte.edu.jm/cms/Portals/0/JBTE/Syllabi/EC%20syllabus/MUSIC%20MU%20100%20EC.mpu.pdf>
- Joint Board of Teacher Education. (2013). Quality assurance responsibility. Retrieved from <http://www.jbte.edu.jm/cms/Home.aspx>
- Kinney, D. (2009). Internal consistency of performance evaluations as a function of music expertise and excerpt familiarity. *Journal of Research in Music Education* 56(4), 322–337. doi:10.1177/0022429408328934
- Lien J. L., & Humphreys, J. T. (2001). Relationships among selected variables in the South Dakota all-state band auditions. *Journal of Research in Music Education*, 49(2), 146–155. doi:10.2307/3345866
- Ross, V. (2009). External music examiners: Micro-macro tasks in quality assurance practices. *Music Education Research*, 11(4), 473–484. doi:10.1080/14613800903390774
- Ryan, C., & Costa-Giomi, E. (2004). Attractiveness bias in the investigation of young pianist's performances. *Journal of Research in Music Education*, 52(2), 141–154.
- Thompson, S., & Williams, A. (2003). Evaluating evaluation: Musical performance assessment as a research tool. *Music Perception*, 21(1), 21–41. doi:10.1525/mp.2003.21.1.21
- Tucker, J. (1995). Music in Jamaican schools and teachers' colleges: Retrospect and prospect. *Education and Research in the Caribbean*, 1(1), 52–67.
- Virtue, E. (2013, May 5). Lessons for women – Females continue to dominate the number of teachers in training. *The Gleaner*. Retrieved from <http://jamaica-gleaner.com/gleaner/20130505/lead/lead1.html>