DURMS
Durham Undergraduate
Research in Music & Science

Do Programme Notes Influence Listening Experiences?

Menglin Luo, Tiffany Mok, Alicia Sher, Guoyu Wang, Zoe Solomon
Durham University

ABSTRACT

Programme notes are vital components of classical music concerts. The current study, which was replicated from Fischinger et al. (2018), investigated the impact of programme notes on music perception and appreciation, with an exploration of the helpfulness of programme notes and considerations of differences between cultural backgrounds. Both Non-western and Western participants were recruited, and data was collected using Qualtrics online questionnaire. Results differed from Fischinger et al. (2018), where the differences in Perceived Musical Characteristics and Liking between experimental conditions were not statistically significant (H1). There was no statistically significant difference between the two cultural groups regarding the impact of programme notes on music perception (H2). Moreover, most participants did not find programme notes very helpful. These findings suggested that the veracity of Fischinger et al. (2018) could have been challenged in a transcultural context.

1. INTRODUCTION

Providing programme notes to the audience at classical music performances is a custom that dates to the 18th century (Blom et al., 2016). The programme notes, which contain details about the background and composer of the musical pieces as well as the compositional ideas (such as structure, meaning, or aesthetic values) that underlie the work, can serve as a guide for audience members (Bennett & Blom, 2014; Fischinger et al., 2018). Contextual information has been discovered to be crucial to how music is perceived since it can influence the emotions that music induces, which in turn shapes the listening experience (Vuoskoski & Eerola, 2013). Additionally, providing additional context about the music prior to the concert programming has been shown to help new classical music audiences better understand the concert and the music (Dobson, 2010). The enjoyment of the music by the audience is essential because it could encourage people to attend concerts and develop an interest in classical music (Thompson, 2006). Due to the importance of the contextual information provided by the programme notes, investigating the impact of programme notes on listening experiences would have practical implications for concert planning and interaction with audiences.

The replication study. This study is replicated from "If it's Mozart, it must be good? The influence of textual information and age on musical appreciation" by Fischinger et al. (2018). It aimed to investigate whether and how the perception and appreciation of a music piece might be influenced by information provided through programme notes, with consideration of influences brought by the age and music expertise of participants. Fischinger et al. (2018) applied a framing paradigm to the listening experiment, where the writing mode and the authorship were manipulated through a 2 × 2 between-factorial design (Composer: Mozart vs Mysliveček and Writing Mode: Analytical vs Expressive). Musical appreciation was measured by quantitative data of Liking (scale 1-6), and 28 semantic differentials for Perceived Musical Characteristics (scale 1-7). These 28 semantic differentials are attributed to four factors: Affective Character, Quality of Artistic Expressivity, Vitality, and Emotional Expressivity. Participants self-reported their familiarity with the music piece and composer, and their music expertise was measured by Gold-MSI scales. Regarding the writing mode of the programme notes, Fischinger et al. (2018) found that participants who read expressive programme notes rated higher for Liking and Affective Character of Perceived Musical Characteristics. For differences in age, younger participants were more likely to be affected by the prestige of the composer.

Extensions and modifications. Replication of Fischinger et al. (2018) in this study is beneficial in assessing the reliability and confidence of the main findings and increasing the sheer number of observations regarding the topic of programme notes (Frieler et al., 2013). This study mainly followed the aims and procedures of Fischinger et al. (2018) and further addressed several limitations of the original study. Firstly, Fischinger et al. (2018) only recruited participants (N = 170) from Germany, and many were interested in classical music and had a higher average Gold-MSI score, hence the findings were constrained to the cultural context. To determine whether listening experiences may alter depending on cultural background, this study recruited participants from a variety of cultural backgrounds. Secondly, the high familiarity with classical music of the German participants might be a

causally influential factor in the listening experiences (Frieler et al., 2013). Contrarily, due to its cultural distinctiveness, classical music would be less familiar to many of the Non-western participants in this study. This made it possible for the study to evaluate the trans-cultural validity of the key findings in Fischinger et al. (2018). Finally, Fischinger et al. did not look at the significance of contextual information in the listening experience (2018). Thus, a control group without any programme notes was introduced in this study. The questionnaire also asked participants about how helpful they found the programme notes.

Research questions, aims and objectives. The research question of this study is "Do programme notes influence listening experiences?", which aims to investigate whether the presence of programme notes, and their writing modes influence the perception and appreciation of a musical work, with the active consideration of variation between Western and Non-western cultural backgrounds as a secondary aim.

Hypotheses. The main hypothesis (H1) predicts that programme notes can influence the listening experience, where (a) the listening experience would differ between participants with programme notes and those without (Bennett & Ginsborg, 2017), and (b) participants with Expressive programme notes would rate higher in Liking and Affective Character than those with Analytical programme notes (Fischinger et al., 2018). Considering that Non-western participants are culturally less familiar with classical music, the secondary hypothesis (H2) predicts that there would be a difference in listening experience and responses to programme notes between Western and Non-western participants. In addition, this study explored if the presence of programme notes is helpful for participants, and whether the listening experience is related to the ratings of Helpfulness.

2. METHODS

Design. This empirical study mainly replicated the procedure of Fischinger et al. (2018). To investigate the primary aim of the impact of programme notes on listening experiences, a between-participant experiment was employed in the form of an online survey. Unlike Fischinger et al. (2018) which manipulated both writing modes and authorship, this study focused on writing modes as the independent variable. Each participant was randomly allocated to a single condition: one controlled group without programme notes and two experimental groups with either Analytical or Expressive programme notes. The dependent variable of listening experience was measured quantitatively by 10 semantic differentials of Perceived Musical Characteristics extracted from Fischinger et al. (2018) and the Liking of the piece. The study also includes categorical variables of age, gender, cultural background, and music expertise, in which the variations in the cultural background on listening experience were explored as a secondary aim.

Participants. Participants were recruited through convenience sampling, anyone above 18 years old and fluent in English was eligible to participate. Links and QR codes of the survey were sent to friends and classmates, shared on social media, and via email to college students. 69 participants took part in this experiment, and all fully completed the survey. The age of participants ranges from 18 to 64 (M = 23.2), while 84% (n = 58) of the participants were aged between 18-24. More than half of the participants were female (n = 44), 20 were male, and 5 were of other genders. Since cultural background is hard to define, this study asked for the country where participants have spent most of their life, as an indication of cultural background. 17 countries were identified, of which UK (n = 34) and China (n = 18, including Hong Kong SAR.) were the most common. Countries were then classified as Western and Non-western, where 46 participants identified as Western and 23 as Non-western.

The music expertise of participants was measured using Ollen-MSI (MARCS Institute, 2006). 16 participants were considered musicians (OMSI > 500). As shown in Figure 1, Western participants demonstrated slightly higher music expertise (M = 319) than Non-western (M = 230). 10 participants self-reported being very familiar (rating = 7) with classical music, while 7 were not familiar at all (rating = 1). Western participants demonstrated a higher level of familiarity with classical music (Figure 2). 21 participants self-reported that they hardly ever listen to classical music. A positive Spearman's rank correlation (r_s = 0.66) was found between self-reported frequency of listening to classical music and familiarity with classical music (Table 1). Regarding the familiarity with the music stimuli and the composer, only 6 participants had heard the piece before, while only 2 indicated that they were "very familiar" with the composer.

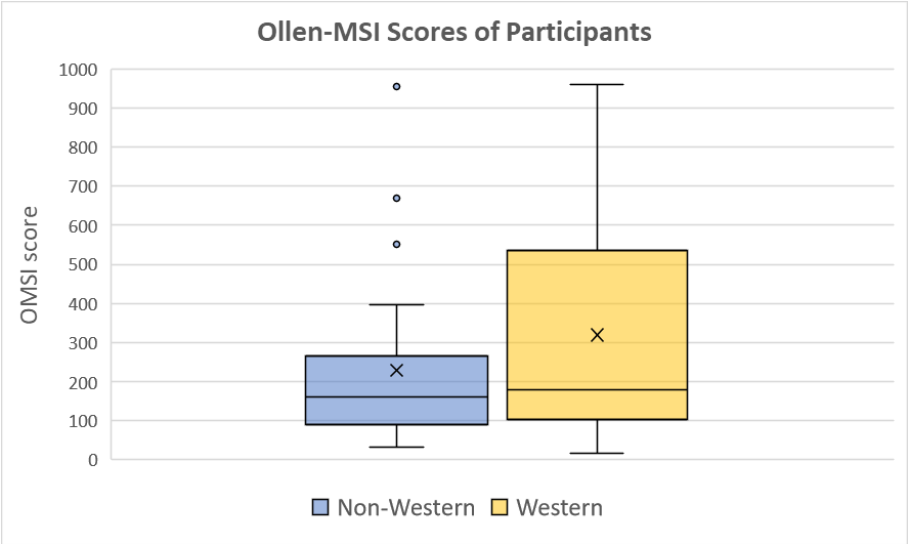

Figure 1. The Ollen-MSI scores of Non-western and Western participants

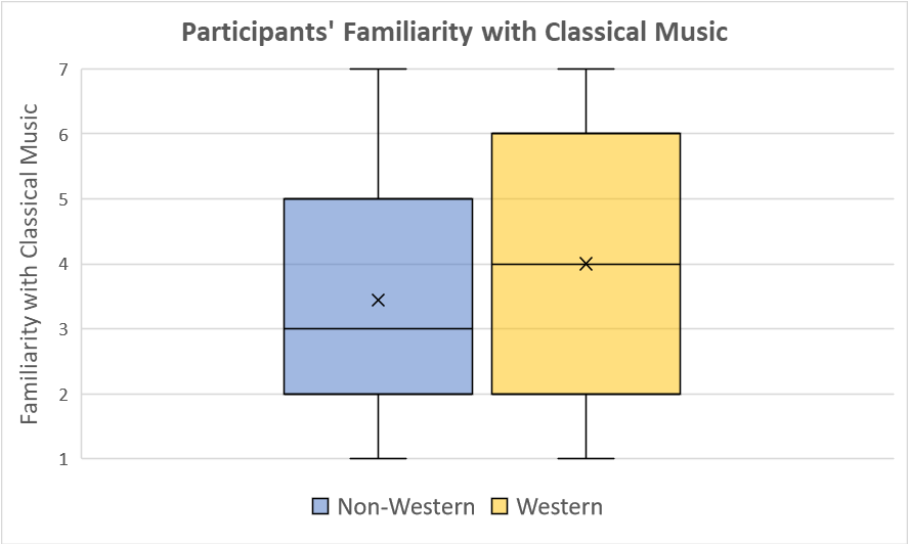

Figure 2. Self-reported familiarity with classical music of Non-western and Western participants

Table 1. The crosstabulation of participants' self-reported frequency of listening to classical music and familiarity with classical music.

Count	**Familiarity with Classical Music (1= Not at all, 7=Very familiar)**							
Frequency of Listening to Classical Music	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Grand Total**
1-3 times per week			4	5	6	2	1	18
A few times a month		3	7	2	3	3	2	20
Daily/5-7 times per week		1	1			1	7	10
Hardly ever	7	9	2	1	2			21
Grand Total	**7**	**13**	**14**	**8**	**11**	**6**	**10**	**69**

Materials/Stimuli. For the convenience of data collection and recruiting participants beyond the geographical constraints of Durham, the experiment used Qualtrics online questionnaire. To draw comparisons with Fischinger et al. (2018), this study used the same programme notes and the same music stimuli, which was the 1st movement "Spirituosa" of Sinfonia in E-flat major composed by Josef Mysliveček in the 18th century (Concerto Köln, 2018). Moreover, the bipolar ratings for semantic differentials (1-7) and Liking (1 = not at all, 6 = very much) used in this study were identical to the original study (Figure 3). 6 semantic differentials belong to the Affective Character factor, which was picked due to its statistical significance ($p < .001$) in writing modes in Fischinger et al. (2018). The other 4 semantic differentials belong to the Emotional Expressivity factor, which might have the potential to have statistical significance ($p = .11$). 10 semantic differentials were randomised in Qualtrics for each participant to eliminate possible order effects (Price et al., 2017).

Figure 3. 10 Semantic differentials of Perceived Musical Characteristics used in the Qualtrics survey, extracted from Fischinger et al. (2018)

Procedure. Upon reading over the information sheet and giving consent to participate in this study, participants were first asked demographic questions about age, gender, cultural background, music expertise, familiarity, and frequency of listening to classical music. Then, participants were randomly allocated to one of the three conditions. All participants were given the title and composer of the music. Participants with programme notes were instructed to read the whole programme notes before listening to the music. A timer was set for the music stimuli, ensuring all participants listened to the whole piece before answering the questions. After listening to the piece, participants were instructed to rate the 10 semantic differentials of Perceived Musical Characteristics and Liking, self-reporting whether they had heard the piece before, and their familiarity with the composer. For the two programme notes conditions, an additional question regarding the helpfulness of the programme notes (bipolar, 1 = very unhelpful; 7 = very helpful) was asked.

Statistical analyses. Statistical analyses were performed in Excel. The factor scores of Affective Character and Emotional Expressivity were calculated from the means of the semantic differentials in each factor group. Single-factor Analysis of Variance (ANOVA) on Affective Character, Emotional Expressivity and Liking were performed across the three conditions, then between Western and Non-western groups. T-tests were applied to the two programme notes conditions on the helpfulness of programme notes, followed by ANOVA analysis between the Western and Non-western groups.

3. RESULTS

Listening experiences. To investigate the impact of programme notes on the listening experience (H1a), the mean ratings of Liking, Affective Character, and Emotional Expressivity were calculated and compared across the three conditions (Figure 4). The mean ratings of these three categories were higher in the No Programme Notes condition than in the two conditions with programme notes. Among the means of Liking, Affective Character, and

Emotional Expressivity of the three conditions, ANOVA results indicated no statistical significance ($p < .05$) (Table 2). To assess the effect of writing modes on the listening experience (H1b), t-tests were conducted to compare the Analytical and the Expressive programme notes conditions, but no statistical significance was found in Affective Character ($t(40) = 0.05$, $p = .96$), Emotional Expressivity ($t(39) = -0.01$, $p = .99$) and Liking ($t(43) = 0.40$, $p = .69$).

Table 2. Mean values (*SD* in brackets) of Liking, Affective Character, Emotional Expressivity, and results of one-way ANOVAs across the three conditions

	No Programme Notes (*n* = 23)	Analytical Programme Notes (*n* = 24)	Expressive Programme Notes (*n* = 22)	Total (*n* = 69)	ANOVA Results
Liking of the piece	4.43 (1.08)	4.21 (1.32)	4.05 (1.43)	4.23 (1.27)	$F(2,66) = 0.52$, $p = .59$
Affective Character	4.93 (0.93)	4.59 (0.86)	4.58 (1.08)	4.70 (0.96)	$F(2,66) = 1.03$, $p = .36$
Emotional Expressivity	4.80 (0.73)	4.67 (0.78)	4.67 (1.01)	4.71 (0.84)	$F(2,66) = 0.20$, $p = .82$.

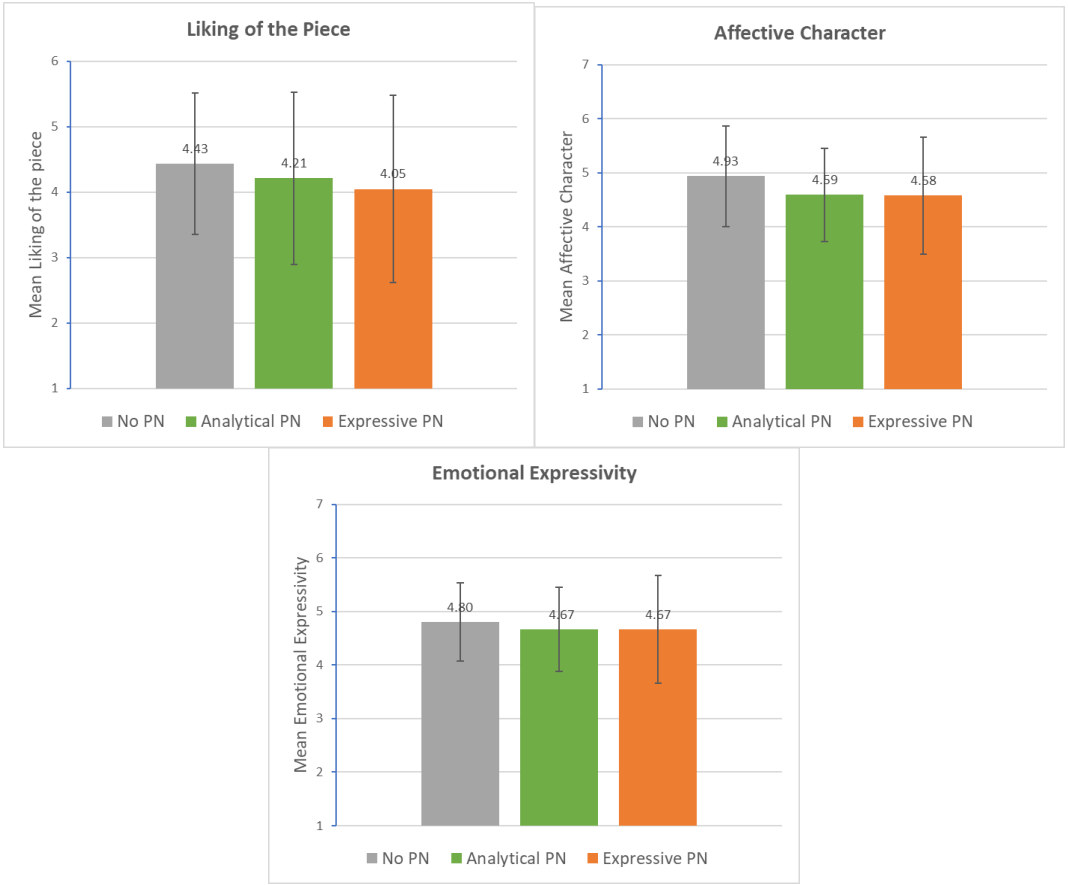

Figure 4. Mean values of Liking, Affective Character, and Emotional Expressivity of the three conditions

Differences in listening experience between Western and Non-western participants in each condition were examined (H2). Apart from the means of Liking and Emotional Expressivity in the No programme notes condition, all other means were higher amongst Western participants regardless of which condition they were in (Figure 5). Results of ANOVA show no statistical significance in Liking, Affective Character, and Emotional Expressivity between Western and Non-Western participants (Table 3).

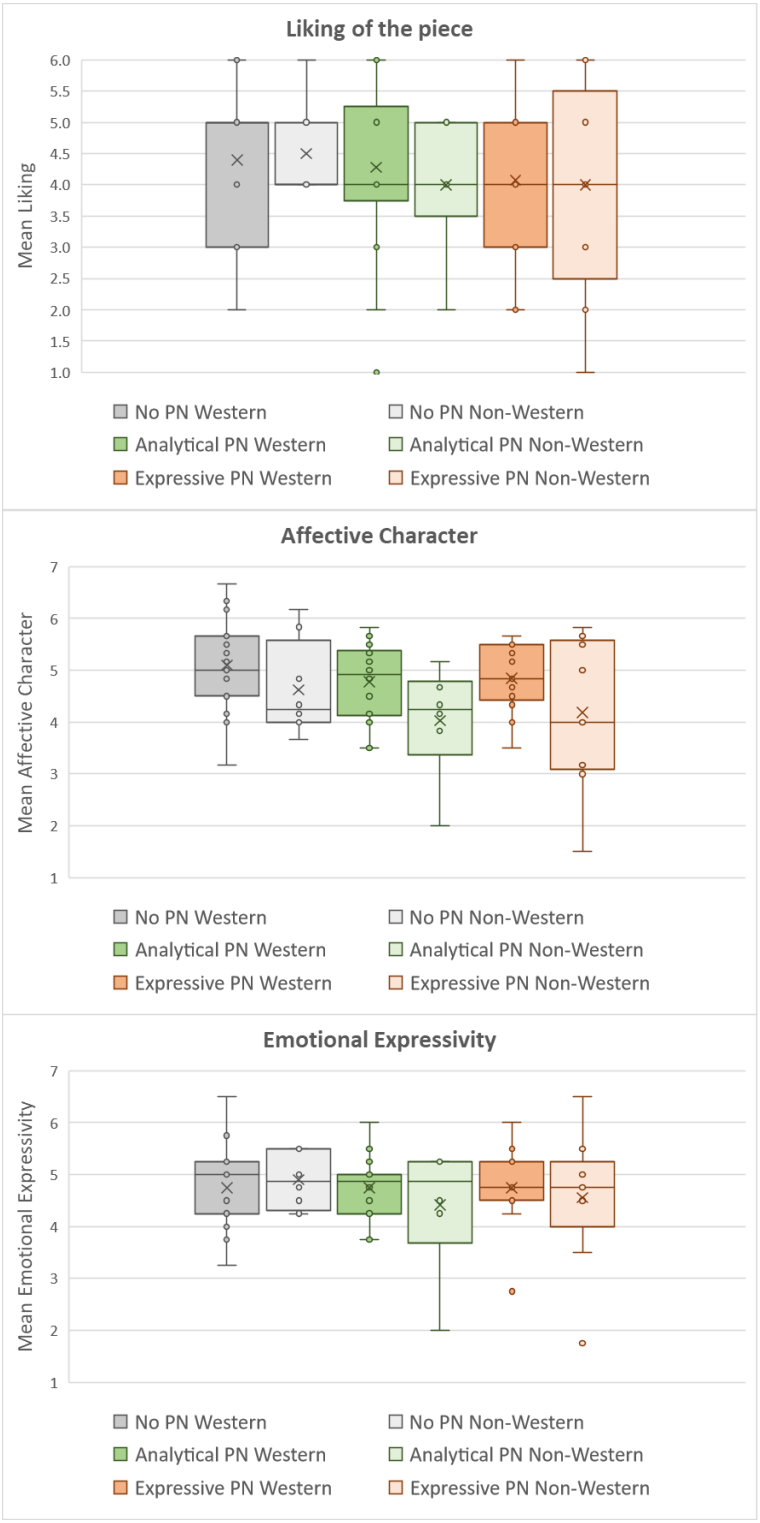

Figure 5. Boxplots summarising the ratings of Liking, Affective Character, and Emotional Expressivity of Western and Non-western participants in the three conditions. Mean values are indicated by the "×" symbol on the plots.

Table 3. Mean values (*SD* in brackets) of Liking, Affective Character, Emotional Expressivity, and results of one-way ANOVAs across Western and Non-western participants in all three conditions

	No Programme Notes		Analytical Programme Notes		Expressive Programme Notes		ANOVA Results
	Western (*n* = 15)	Non-Western (*n* = 8)	Western (*n* = 18)	Non-Western (*n* = 6)	Western (*n* = 13)	Non-Western (*n* = 9)	
Liking of the piece	4.40 (0.76)	4.50 (1.24)	4.28 (1.41)	4.00 (1.10)	4.08 (1.26)	4.00 (1.73)	$F(5,63) = 0.25$, $p = .94$
Affective Character	5.10 (0.92)	4.63 (0.92)	4.78 (0.71)	4.03 (1.09)	4.85 (0.67)	4.19 (1.46)	$F(5,63) = 1.82$, $p = .12$
Emotional Expressivity	4.75 (0.83)	4.91 (0.55)	4.75 (0.58)	4.42 (1.26)	4.75 (0.78)	4.56 (1.32)	$F(5,63) = 0.30$, $p = .91$

Correlations (*r*) between factor scores, and between factor scores and Liking were calculated, to explore if these measurements of listening experiences correlate with each other (Table 4). According to Akoglu (2018)'s interpretation of the strength of the relationship between variables, moderate ($r > 0.4$) and strong ($r > 0.7$) were found between Affective Character and Emotional Expressivity across all three conditions (Figure 6), as well as between Affective Character and Liking in the No programme notes condition.

Table 4. Pearson's correlation coefficient (*r*) of Affective Character vs. Emotional Expressivity, Affective Character vs. Liking, and Emotional Expressivity vs. Liking.

	No Programme Notes (*n* = 23)	Analytical Programme Notes (*n* = 24)	Expressive Programme Notes (*n* = 22)
Affective Character vs. Emotional Expressivity	0.62	0.68	0.80
Affective Character vs. Liking	0.48	0.20	0.07
Emotional Expressivity vs. Liking	0.33	-0.07	-0.05

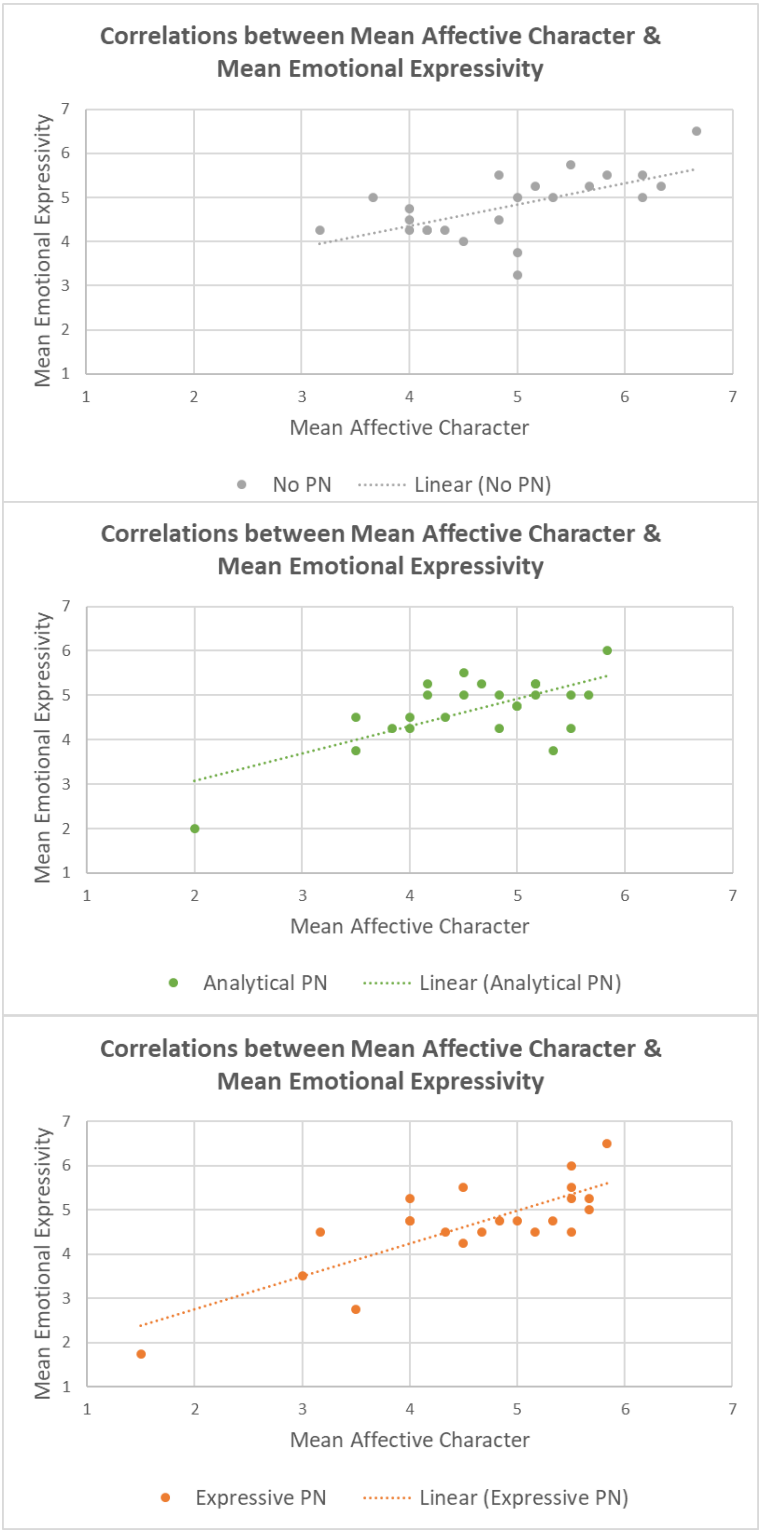

Figure 6. The scattered plots demonstrating correlations between Affective Character and Emotional Expressivity across the three conditions

Helpfulness of programme notes. The mean ratings of Helpfulness were similar between the Analytical ($M = 4.75$, $SD = 1.42$) and the Expressive programme notes condition ($M = 4.77$, $SD = 1.74$) (Figure 7). The result of the t-test comparing the mean Helpfulness ratings between the two conditions suggests no statistical significance ($t(41) = -0.05$, $p = .96$). No obvious correlations were found between Helpfulness and Affective Character, or Emotional Expressivity (Appendix 1). A moderate correlation ($r = 0.45$) in the Analytical programme notes condition and a strong correlation ($r = 0.77$) in the Expressive programme notes condition were found between Liking and Helpfulness. In both conditions with programme notes, the ratings of Helpfulness were higher in Non-western groups ($M_A = 5.50$, $SD_A = 1.05$; $M_E = 4.89$, $SD_E = 2.03$) than in Western groups ($M_A = 4.50$, $SD_A = 1.47$; $M_E = 4.69$, $SD_E = 1.60$) (Figure 8). To assess if the mean ratings of Helpfulness differ between the four groups, ANOVA was performed, which result indicated no statistical significance ($F(3,42) = 0.62$, $p = .60$).

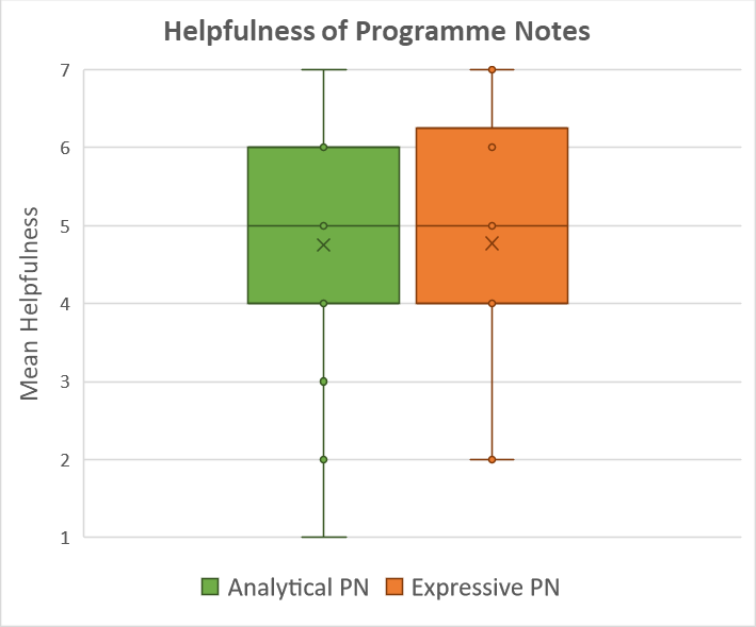

Figure 7. Boxplots summarising the mean ratings of the helpfulness of programme notes in the two conditions with programme notes. Mean values are indicated by the "×" symbol on the plot.

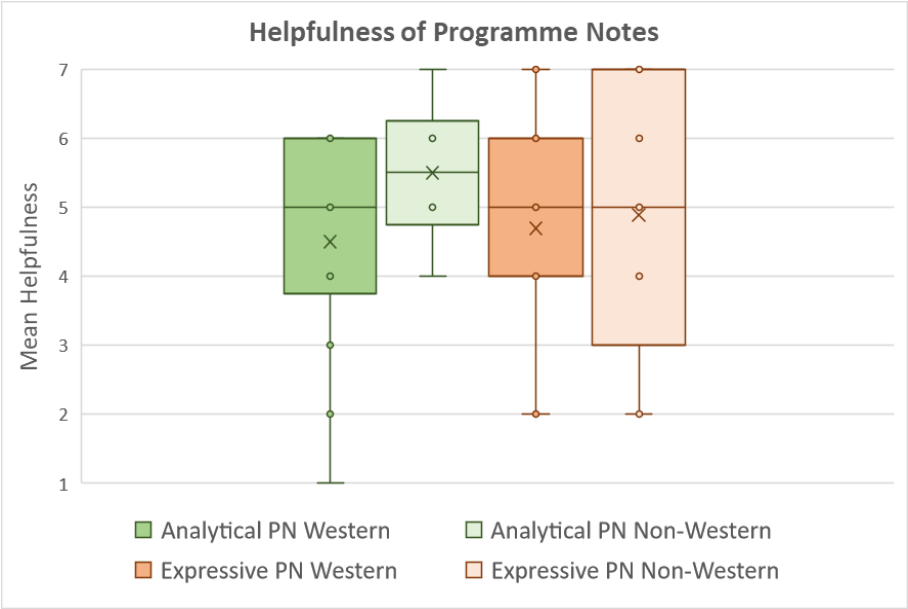

Figure 8. Boxplots summarising the mean ratings of helpfulness of programme notes of Western and Non-western participants in the two conditions with programme notes. Mean values are indicated by the "×" symbol on the plot.

4. DISCUSSION

Impact of programme notes on listening experiences (H1). This study predicted that the listening experiences would differ between participants with and without programme notes (H1a). Although the mean ratings of Liking and Perceived Musical Characteristics were higher in the No programme notes condition, ANOVA indicated no statistically significant differences between the three conditions. Regarding the impact of writing modes on listening experiences (H1b), unlike in Fischinger et al. (2018) where the participants with Expressive programme notes rated higher for Liking and Affective Character, in this study, not only minimal differences were found in the means of Liking, Affective Character, and Emotional Expressivity between the two writing modes, but the t-tests also indicated no statistical significance between the two conditions.

One possibility that the ANOVAs and t-tests were not statistically significant is due to the high standard deviation of the ratings, where the data were widely spread. Therefore, the means were not representative of the dataset. Additionally, the *F*-statistics from ANOVA on Liking, Affective Character, and Emotional Expressivity were small (Table 3), suggesting that there was greater variation within each condition than between conditions (Kim, 2017). As a result, the presence of programme notes, and the writing modes were unlikely to have an impact on the listening experiences of participants.

Despite none of the measurements of listening experience demonstrated statistical significance, the *p*-value of Affective Character ($p = .12$) was much lower than Liking and Emotional Expressivity, which is similar to Fischinger et al. (2018). Therefore, Affective Character might have the potential of being affected by programme notes, which can be explored in future research with a larger sample size. Moreover, positive correlations between Affective Character and Emotional Expressivity were found in all three conditions, which is comparable to Fischinger et al. (2018). Nevertheless, positive correlations between factor scores and Liking were also presented in Fischinger et al. (2018) but this was not the case in this study.

Cultural differences in listening experiences of different programme notes (H2). Differences in listening experiences were expected (H2), but the results were not supportive of this hypothesis. Although the mean ratings of Liking and Perceived Musical Characteristics were generally higher in the Western group, ANOVAs demonstrated no statistical significance between the means of different cultural groups. In addition to the large standard deviations and small *F*-statistics, the sample sizes of Western and Non-western participants were not balanced in each condition, which could have weakened the statistical power.

Interestingly, though with no statistical significance, Affective Character again demonstrated a lower *p*-value than Liking and Emotional Expressivity between Western and Non-western groups, and Western participants reported higher ratings in Affective Character. Considering that most Non-western participants were Chinese, the potential differences in Affective Character might be related to the mode of expression in different cultures, where Westerners are more direct in expressing opinions, whereas Chinese are more implicit (Wang et al., 2021). A larger and more balanced sample may give better insight into the differences in cultural background on the ratings of Affective Character.

Helpfulness of programme notes. The helpfulness of the programme notes was exploratory; thus, no directional hypothesis was made. The mean ratings of Helpfulness of the two programme notes conditions were very close, all had large standard deviations, and the *t*-test was not statistically significant. Additionally, since the question regarding Helpfulness used a bipolar scale, the rating of 4 refers to neither helpful nor helpful. Hence, the low mean ratings for Helpfulness indicate that most participants found neither type of programme notes to be obviously helpful. Relating to the results of Liking of the piece of this study, the unhelpfulness of programme notes might explain the lower ratings of Liking found in the two conditions with programme notes than in the No programme notes condition. This explanation is also in line with Margulis (2010), which suggested that reading programme notes before listening to music reduces the enjoyment of music.

Non-western participants reported the programme notes were more helpful than Western participants, but no statistical significance was found when comparing the means, possibly due to large standard deviations. Nevertheless, in the Analytical programme notes condition, the distinction between Non-western and Western participants is greater, thus demonstrating an interest in future research. Bennett & Ginsborg (2017) noticed that more experienced listeners in classical music are more likely to reject the information on programme notes, in favour of their own interpretations. Since the Non-western participants in this study were generally less familiar with classical music, it is possible that they would find analytical programme notes more helpful than Westerners because it provided a structure to guide them through the piece.

Evaluation. The strengths of this study were recruiting participants with diverse cultural backgrounds, and exploring the helpfulness of programme notes, which widened the discussion regarding the impact of programme notes demonstrated in Fischinger et al. (2018). Several limitations were also identified. Firstly, compared with Fischinger et al. (2018), the sample size of this study was smaller. Although Qualtrics could allocate participants to a random condition in a balanced way, it could not allocate participants based on cultural backgrounds. A larger sample size and more equal distribution of Western and Non-western participants would improve the statistical power. Secondly, the 10 semantic differentials were object-oriented and bipolar (e.g., heavy - light), but are not indicative of the intensity of the Perceived Musical Characteristics, which could be considered in future research. Thirdly, the selected music stimuli was only a short movement from the Classical Era, where musical features such as dynamics, pitch range and structure are much simpler than in later periods. Considering the differences in music perception in different eras, the perceptions would differ in more complex pieces (Kelly et al., 2021). Therefore, the findings of this study should not be generalised to classical music as a whole.

More importantly, data were collected in an experimental setting, where the participants were instructed to read the whole programme notes before listening to the piece. Nevertheless, in reality, audiences would have different preferences on when to read the programme notes and be selective on the information provided. Future research can consider taking a real-life approach and using qualitative approach such as interviewing the audience regarding their habits in reading programme notes.

Conclusions and implications. This study expanded the investigation of the impact of programme notes on different cultural groups. In general, no obvious impact of programme notes and writing modes was found on the listening experience. Programme notes were also not very helpful, as reported by participants. Regarding cultural backgrounds, the t-tests and ANOVAs were not statistically significant, but Affective Character and helpfulness of programme notes implied potential relations to cultural backgrounds, which can be further explored in larger samples. This study showed different results from Fischinger et al. (2018), suggesting that the findings in the original study might be subjective to cultural contexts. Finally, this study highlighted the importance of considering cultural backgrounds in the research of programme notes, as they have real-life implications when presenting classical music to a global audience.

REFERENCES

Concerto Köln. (2018). Mysliveček: Symphony In E Flat For 2 Horns, 2 Oboes And Strings - 1. Spirituosa [YouTube Video]. In *YouTube*. https://youtu.be/f0jnpJuPQb0

Akoglu, H. (2018). User's Guide to Correlation Coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. National Library of Medicine. https://doi.org/10.1016/j.tjem.2018.08.001

Bennett, D., & Blom, D. (2014). The program note as creative knowledge and skills: Shaping a collaborative interpretation of newly composed music. In G. Carruthers (Ed.), *Proceedings of the 20th International Seminar of the ISME Commission on the Education of the Professional Musician* (pp. 173–180). International Society for Music Education. https://www.researchgate.net/publication/268507490_The_program_note_as_creative_knowledge_and_skills_shaping_a_collaborative_interpretation_of_newly_composed_music

Bennett, D., & Ginsborg, J. (2017). Audience reactions to the program notes of unfamiliar music. *Psychology of Music*, *46*(4), 588–605. SAGE Journal. https://doi.org/10.1177/0305735617721339

Blom, D. M., Bennett, D., & Stevenson, I. (2016). The Composer's Program Note for Newly Written Classical Music: Content and Intentions. *Frontiers in Psychology*, *7*. Frontiers. https://doi.org/10.3389/fpsyg.2016.01707

Dobson, M. C. (2010). New Audiences for Classical Music: The Experiences of Non-attenders at Live Orchestral Concerts. *Journal of New Music Research*, *39*(2), 111–124. https://doi.org/10.1080/09298215.2010.489643

Fischinger, T., Kaufmann, M., & Schlotz, W. (2018). If it's Mozart, it must be good? The influence of textual information and age on musical appreciation. *Psychology of Music*, *48*(4), 030573561881221. SAGE Journal. https://doi.org/10.1177/0305735618812216

Frieler, K., Müllensiefen, D., Fischinger, T., Schlemmer, K., Jakubowski, K., & Lothwesen, K. (2013). Replication in music psychology. *Musicae Scientiae*, *17*(3), 265–276. https://doi.org/10.1177/1029864913495404

Kelly, B. O., Anderson, C. J., & Schutz, M. (2021). Exploring Changes in the Emotional Classification of Music between Eras. *Auditory Perception & Cognition*, *4*(1-2), 121–131. Taylor & Francis. https://doi.org/10.1080/25742442.2021.1988422

Kim, T. K. (2017). Understanding one-way ANOVA using conceptual figures. *Korean Journal of Anesthesiology*, *70*(1), 22. NCBI. https://doi.org/10.4097/kjae.2017.70.1.22

MARCS Institute. (2006). *Ollen Musical Sophistication Index (OMSI)*. Marcs-Survey.uws.edu.au; MARCS Institute. http://marcs-survey.uws.edu.au/OMSI/

Margulis, E. H. (2010). When program notes don't help: Music descriptions and enjoyment. *Psychology of Music*, *38*(3), 285–302. SAGE Journal. https://doi.org/10.1177/0305735609351921

Price, P. C., Jhangiani, R. S., Chiang, I-Chant. A., Leighton, D. C., & Cuttler, C. (2017). Experimental Research. In *Research Methods in Psychology*. Press Books. https://opentext.wsu.edu/carriecuttler/

Thompson, S. (2006). Audience responses to a live orchestral concert. *Musicae Scientiae*, *10*(2), 215–244. https://doi.org/10.1177/102986490601000204

Vuoskoski, J. K., & Eerola, T. (2013). Extramusical information contributes to emotions induced by music. *Psychology of Music*, *43*(2), 262–274. SAGE Journals. https://doi.org/10.1177/0305735613502373

Wang, X., Wei, Y., Heng, L., & McAdams, S. (2021). A Cross-Cultural Analysis of the Influence of Timbre on Affect Perception in Western Classical Music and Chinese Music Traditions. *Frontiers in Psychology*, *12*. Frontiers. https://doi.org/10.3389/fpsyg.2021.732865

APPENDICES

Appendix 1. Pearson's correlation coefficient (r) between factor scores, Liking and helpfulness of programme notes across the three conditions.

	No Programme Notes (*n* = 23)	**Analytical Programme Notes (*n* = 24)**	**Expressive Programme Notes (*n* = 22)**
Affective Character vs. Emotional Expressivity	0.62	0.68	0.80
Affective Character vs. Liking	0.48	0.20	0.07
Emotional Expressivity vs. Liking	0.33	-0.07	-0.05
Liking vs. Helpfulness of PN	N/A	0.45	0.77
Affective Character vs. Helpfulness of PN	N/A	-0.08	0.19
Emotional Expressivity vs. Helpfulness of PN	N/A	=0.09	0.02