DOI: 10.1111/nyas.14944

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

REVIEW

Quality standards and recommendations for research in music and neuroplasticity

Efthymios Papatzikis[1] | Maria Agapaki[1] | Rosari Naveena Selvan[2,3] | Varun Pandey[4] | Fathima Zeba[5]

[1]Department of Early Childhood Education and Care, Oslo Metropolitan University, Oslo, Norway

[2]Institute for Physics 3 - Biophysics and Bernstein Center for Computational Neuroscience (BCCN), University of Göttingen, Göttingen, Germany

[3]Department of Psychology, University of Münster, Münster, Germany

[4]NeuralORCH A.I. LLP, Delhi, India

[5]School of Humanities and Social Sciences, Manipal Academy of Higher Education Dubai, Dubai, United Arab Emirates

Correspondence
Efthymios Papatzikis, Department of Early Childhood Education and Care, Oslo Metropolitan University, Pilestredet 42, Oslo 0167, Norway.
Email: efthymio@oslomet.no

Abstract

Research on how music influences brain plasticity has gained momentum in recent years. Considering, however, the nonuniform methodological standards implemented, the findings end up being nonreplicable and less generalizable. To address the need for a standardized baseline of research quality, we gathered all the studies in the music and neuroplasticity field in 2019 and appraised their methodological rigor systematically and critically. The aim was to provide a preliminary and, at the minimum, acceptable quality threshold—and, *ipso facto*, suggested recommendations—whereupon further discussion and development may take place. Quality appraisal was performed on 89 articles by three independent raters, following a standardized scoring system. The raters' scoring was cross-referenced following an inter-rater reliability measure, and further studied by performing multiple ratings comparisons and matrix analyses. The results for methodological quality were at a quite good level (quantitative articles: mean = 0.737, SD = 0.084; qualitative articles: mean = 0.677, SD = 0.144), following a moderate but statistically significant level of agreement between the raters ($W = 0.44$, $\chi^2 = 117.249$, $p = 0.020$). We conclude that the standards for implementation and reporting are of high quality; however, certain improvements are needed to reach the stringent levels presumed for such an influential interdisciplinary scientific field.

KEYWORDS
neuroplasticity and music, neuroscience and music, neuroscience research methodology, QualSyst research evaluation system, research recommendations, research standards

INTRODUCTION

The brain is a unique organ that demonstrates widespread variability within and across populations, with differences being described in structural and functional terms. One aspect of brain differences that has gained interest in recent years is the study of neuroplasticity. Neuroplasticity refers to the dynamic physiological changes that occur in the brain resulting from the organism's interaction with the environment. Plasticity of the brain can be best defined as "the brain's ability to create adaptive changes in morphological and network neuronal structure and function of the nervous system, which includes changes in neuronal connectivity, neurogenesis and neurochemical changes."[1] Neuroplasticity, viewed as sensitivity to sensory information, can lead to positive and negative effects on the brain. Here, we define "positive" as (1) change that is beneficial or related to a difference in skill, learning, or ability, such as the brain's response to music training;[2] and (2) the rehabilitative effects of music on the brain after injury or neurodegenerative diseases.[3] Negative neuroplastic effects are defined as a change that happens at the molecular, functional, and structural levels in different pathophysiological conditions, such as stress or depression.[4,5]

Music—long considered only as an aesthetically pleasurable experience[6]—has been much discussed in terms of phylo-ontogenetic neuroplasticity in recent years.[7–11] This fact may perhaps be attributed to two major reasons. The first one is that music can be experienced without its main perceptual apparatus—the ears—being functional. Because of that, music's usage in the particular domain of inquiry favors research designs in terms of implementation and utilitarianism. Most other aesthetic-sensory–driven, creative human activities require a dedicated sensory organ to achieve similar outcomes, such as eyes for visual art. For music, on the other hand, an individual can still have a valuable experience—and consequently induce neuroplastic changes in the brain—after perceiving vibrations through their peripheral nervous system, epithelium, and bones. In its simplest form, this is evidenced by fetuses as well as deaf individuals who are able to perceive and respond to music.[12] The second reason is music's challenging learning profile. This challenging learning profile—during bimanual instrument training, for example—seems to provide a multilevel link between microscopic and macroscopic, cortical, and subcortical changes.[13–17]

The great interest in music's relevance to neuroplasticity led us to find links of its underlying neuronal mechanisms with essential adaptive functions, such as emotional communication, social bonding, and mate selection.[18–21] It also led us to find evidence related to "therapeutic" traits or biological and cognitive paths of development[22–27]—evidence that may even suggest far-transfer effects, according to some studies (for a review, see Ref. 28). Based on the exciting nature of this emerging knowledge, many of the aforementioned links and evidence have either created more momentum for evidence-based practices claiming causative changes (e.g., see Ref. 29) or, conversely, have been challenged as being unuseful due to their inability to report causation, when they were actually based on correlative research designs (for a detailed list of these studies, see p. 430 of Ref. 30). This latter viewpoint started to become more evident after specialized investigations showed that neuroscientists in the music field may be prone to believe music's causative role in neuroplastic development and consequently report their research in a looser way when compared to their colleagues in other fields.[30]

Given the broad-based nature of the interdisciplinary studies encompassing the field of music and neuroplasticity, as well as the latter's fairly recent state of existence, it is logical to accept, up to a certain extent, for some reporting failures to arise, either due to occasional fundamental design errors or due to inexperienced experimental implementation of these designs when novice researchers undertake investigations in this field. Difficulties and discrepancies are concepts deeply hardwired in the evolution of science as "transitions to maturity have seldom been…sudden or…unequivocal" in nature.[31] And this could not be expected otherwise, considering the extensive number of simultaneous issues being addressed through multiple research questions and answers at such an early stage,[31] which does not have the precedent of a mature and extensive scholarly framework to rely on.[32] What is not or should not be expected, however, is for the methodological standards applied in the field to be one of the main reasons feeding controversy on the subject of *if* and *how* music has a direct—causative—link to development, and more so to the extended neuroplastic system of our brain.

Considering the wide range of approaches and parameters that are employed by studies linking music and the brain (e.g., electroencephalogram [EEG]/ERP;[33,34] functional magnetic resonance imaging [fMRI];[35] diffusion tensor imaging [DTI];[36] voxel-based morphometry;[37] *in vivo* magnetic resonance morphometry;[17] MEG;[38] musicians vs. nonmusicians;[39] and different age groups[40]) as well as the nonuniform—or even low quality, as described by some—methodological standards implemented[2,41–43] in this particular field of inquiry, we indeed end up with findings that are prone to be less generalizable and frequently nonreplicable. As a result, the inconsistent standards of quality for our methods not only provide more justification for those who are skeptical of music's possible direct link to neuroplasticity, but also make it difficult to support future studies that depend on obtaining outcomes that are comparable, interconnected, and extendable to other domains. In view of this, a need for a unified baseline in research quality arises, as has been stated before.[44,45]

To fill this gap, the current study first gathered and analyzed the available methodological approaches employed by both qualitative and quantitative studies in the music and neuroplasticity field in one calendar year, namely 2019, and then appraised their methodological quality systematically and critically. The aim of this study was not of course to present an absolute baseline of methodological quality and standards, but to merely provide a preliminary, at the minimum, acceptable quality threshold—and, *ipso facto*, suggested recommendations—via systematic exploration, whereupon further discussion and perhaps development may occur for the benefit of the field. The quality appraisal was performed by three independent raters, following a standardized scoring system based on a pairwise inter-rater agreement tool specialized for this kind of inquiry. The raters' scoring was cross-referenced following an inter-rater reliability (IRR) measure, and further analyzed by performing multiple ratings comparisons. In our study, quality was evaluated—as was done by previous studies in similar exploratory cases in other scientific fields[46]—in terms of internal validity, or the extent to which the overall design, conduct, and data analyses minimized errors or biases.

METHODOLOGY

Sample

For our study, 89 articles[47–132] were extracted from the Google Scholar and PubMed databases using the keywords "neuroplasticity," "brain changes," "music," and "sound." Based on these keywords, we included all articles that were published in 2019, written in English, and peer-reviewed. The sample included both quantitative and qualitative studies. The articles were screened to include only topics referring directly or indirectly (as meta-analyses and reviews) to experimental or intervention studies. For our sample, to firmly substantiate the music–neuroplasticity methodological connection, we accepted only those articles (1) including neurophysiological measures of brain regions and networks showing brain activation in response to music; (2) using behavioral assessments of psychological functions in at least two treatment groups, where one group is with musical exposure and one group is without musical exposure; and (3) showing longitudinal structural changes in the brain or behavioral changes as an effect of music. Each article was specifically scrutinized to see if it met at least one of these conditions. The assumption in selecting the articles was that neuroplasticity is either directly observed as structural or functional changes or implied through behavioral changes, such as improved processing speed or executive capacity, that cannot happen without neuronal changes. Research looking at music as a subsidiary of dance-related intervention training or using sounds or pure tones was not included.

Raters' background

The methodological quality of the final 89 articles included in our investigation was rated by three independent raters, designated A, B, and C. The raters pursued their higher studies in India, Germany, and Greece, with an average of 6–8 years of education in the field of psychology and brain sciences. Out of the three, two had no academic or research background in music, while the third had an academic background in music performance and production. All the raters were early-stage researchers in the field of psychology and neuroscience at the time of the rating. The raters were not informed about the aim and background of the study before or during the rating process. They were instructed to read the articles that were included for the study and rate them using the 3-point ordinal scale (0–2 and N/A) developed by Kmet et al.,[133] as explained below.

Rating and statistical analysis

Power of sample and content descriptives

Before starting the methodological quality rating process, we first decided to see if our sample of articles was adequately representative—when compared to the overall population of articles in the specific field—to offer a generalized approach of results. Then, we configured the included articles' content profile. For this, we used two categories: quantitative studies ($n = 81$) and qualitative studies ($n = 8$). Based on this categorization, the articles were analyzed and coded for their content ("content profile" hereafter), aiming for basic comprehension of their main methodological framework. A data extraction protocol was developed to extract relevant data under the following four categories: "methods," "tools used," "number of participating sample," and "research site." The "methods" category referred to the quantitative and qualitative mode of study; the "tools used" referred to the neuroimaging and non-neuroimaging research modalities used in the articles; the "number of participating sample" was coded into six categories, with a class interval of 20; and the "research site" referred to the "Lab," "Mixed," and "Not Lab" categories.

The quality investigation approach

Rating and acquisition

Following the content profile analysis, a quality-of-methodology rating was performed by the three raters on the 89 articles. In order to assess the agreement between raters' responses, we used quantitative and qualitative criteria based on 24 questions (the "checklist items" hereafter; Table 1). These checklist items were derived from the "QualSyst" standardized scoring system, a tool developed by Kmet et al.,[133] based on instruments developed by Cho et al.,[134] Popay et al.,[135] Timmer et al.,[136] and Mays and Pope,[137] to assess the quality of research reports of quantitative and qualitative studies. The quantitative category of ratings in QualSyst contains 14 items, whereas the qualitative category of ratings has 10 items. The checklist items assess the clearness of research questions or objectives, appropriateness of sampling strategy and sample size, experimental design, description of participants' characteristics, the robustness of outcome and exposure variables, analytical methods, estimates of variance, control of confounding factors, statistical analyses, and whether conclusions drawn reflect results reported. The ratings scoring of these checklist items was performed on an agreement scale of 0–2 and N/A (where NA = "Not Applicable," 0 = "Disagree," 1 = "Partially Agree," and 2 = "Completely Agree").

Overall score calculations and reliability analysis

As per the QualSyst manual, two separate scoring formulae were used to calculate methodological quality. Specifically, for the quantitative articles, 14 relevant checklist items were scored based on the degree to which the specific criteria were met ("Completely Agree" = 2, "Partially Agree" = 1, "Disagree" = 0). Checklist items that were not applicable to a specific study design were marked "N/A" and were not included in the calculation of the overall score. An overall score was calculated for each article by summing the total score obtained across the 14 checklist items and dividing it by the maximum possible sum (i.e., 28). The quantitative scoring formula was defined as:

$$\begin{aligned} \textit{Quantitative article overall scoce} = {} & ((\text{number of "completely agree"} \times 2) \\ & + (\text{number of "partially agree"} \times 1))/28 \\ & - (\text{number of "N/A"} \times 2) \end{aligned}$$

The qualitative studies' scores were calculated in a similar fashion, based on the scoring of the 10 relevant checklist items. It was not permissible to assign "N/A" to any of the items, and the overall score for each article was calculated by adding the total score obtained across the 10 items and dividing it by the maximum possible sum (i.e., 20). The qualitative scoring formula was defined as:

$$\begin{aligned} \textit{Quantitative article overall scoce} = {} & ((\text{number of "completely agree"} \times 2) \\ & + (\text{number of}\varepsilon\text{partially agree}\varepsilon \times 1))/20 \end{aligned}$$

TABLE 1 Checklist items for quantitative and qualitative articles

	Quantitative articles	Qualitative articles
1	Question/objective sufficiently described?	Question/objective sufficiently described?
2	Study design evident and appropriate?	Study design evident and appropriate?
3	Method of subject/comparison group selection *or* source of information/input variables described and appropriate?	Context for the study clear?
4	Subject (and comparison group, if applicable) characteristics sufficiently described?	Connection to a theoretical framework/wider body of knowledge?
5	If interventional and random allocation was possible, was it described?	Sampling strategy described, relevant, and justified?
6	If interventional and blinding of investigators was possible, was it reported?	Data collection methods clearly described and systematic?
7	If interventional and blinding of subjects was possible, was it reported?	Data analysis clearly described and systematic?
8	Outcome and (if applicable) exposure measure(s) well defined and robust to measurement/misclassification bias? Means of assessment reported?	Use of verification procedure(s) to establish credibility?
9	Sample size appropriate?	Conclusions supported by the results?
10	Analytic methods described/justified and appropriate?	Reflexivity of the account?
11	Some estimate of variance is reported for the main results?	
12	Controlled for confounding?	
13	Results reported in sufficient detail?	
14	Conclusions supported by the results?	

The final scoring formulae provided a range in scoring from 0 to 1.0 as the maximum achievable score.

Following the calculated overall scores, an IRR analysis was performed in order to assess (1) if the three independent raters reached a consensus between their overall scores given, and (2) the relative strength of consensus. The method used for describing the concordance between the raters was Kendall's (*W*) coefficient of concordance. Kendall's *W* is a nonparametric test used to evaluate the agreement between more than two different raters who are evaluating a given set of *n* objects (in our study, the 89 articles).[138] The values of Kendall's *W* vary between 0 and 1, and the greater the concordance between the raters, the closer the *W* is to 1.[138] Through the IRR analysis, the following hypotheses were tested:

- $\mathbf{H_0}$: judges' ratings for the selected articles in the field of "neuroplasticity and music" are contradictory.
- $\mathbf{H_a}$: judges' ratings for the selected articles in the field of "neuroplasticity and music" are comparable.

To further strengthen our study's reliability, we created a checklist items correlation matrix (Tables S1 and S2) based on the articles' neuroimaging versus non-neuroimaging content profile. This step applied only to the quantitative articles. The analysis was performed on all 14 quantitative checklist items scores to explore if individual quantitative statements have a linear relationship. This exploratory step helped us study even deeper the relationship between the raters' scores. More specifically, we aimed to understand if the checklist items' statement content influenced the raters' rating approach, and if the checklist items' serial order could introduce an unintentional bias in the overall rating process.

Classification of methodological quality

After completing the blind rating process, the calculations of the overall scores, and the IRR measurements, our analysis continued to the appraisal and classification of each article's methodological quality. For this analysis step, we employed the quality-range-of-appraisal tool as defined by Lee et al.[139] According to Kmet et al.,[133] the proposed quality threshold (pass/fail) for quantitative articles should be a strict overall score of 0.75, while for the qualitative ones, it should be a score of 0.55. Lee et al.,[139] on the other hand, offer a range of scores that is more cognizant of the complexity of the methodological design of the articles; this is presented as "strong" (overall score of ≥0.80), "good" (overall score of 0.71–0.79), "adequate" (overall score of 0.50–0.70), and "limited" (overall score of <0.50). Using this more flexible classification, our aim was not to exclude articles from any relevant scientific process, but to discover their strengths and limitations when compared to all the articles published in the field of "neuroplasticity and music" in 2019.

Pairwise and multiple comparisons

The final step of our study included a two-stage analysis that combined our scoring data and quality classification data in two different conditions (i.e., "first condition": *pairwise comparisons* between groups of raters and groups of articles—quantitative/qualitative content; "second condition": *multiple comparisons* between all raters and all articles). This two-stage final step was done to appraise the articles at different comparative levels, and consequently reveal itemized—that is, based on the checklist items—methodological weaknesses. Specifically, this latter more detailed itemized analysis could be used as an impetus for improvement in future research practices in the field of neuroplasticity and music and beyond.

In the "first condition," the three independent raters were categorized into three different groups (**Group 1**: raters A-B; **Group 2**: raters A-C; and **Group 3**: raters B-C). Given that the theoretical framework of Kmet et al.[133] is based on assessing the quality of articles between only two raters, we decided to conduct a ternary repeated simulation to generate more cross-referenced data. Based on this simulation, we evaluated inter-rater agreement for Groups 1, 2, and 3 separately and tried to determine whether differentiated trajectories in overall scores and checklist items exist between the pairs of raters. This first condition provided the requisite cross-referencing step upon which we completed the averaging of scores used in the multiple comparisons "second condition" that followed afterward.

With the second condition, we aimed to decrease the anticipated (due to exogenous background differences; confounding variables) heterogeneity of our first condition findings and obtain a summative methodological quality appraisal for all 89 articles simultaneously, combining all three raters' group scores together. With this more conclusive approach, we would like to obtain averaged data, and consequently, present more homogeneous and generalized results. More specifically, for each quantitative and qualitative article, we calculated the average article's quality score among the three groups of raters by adding the mean of the overall QualSyst's score for each rater's group (**Group 1** + **Group 2** + **Group 3** = Average Scoring). Then, we classified each article using the Lee et al.[137] quality-range-of-appraisal tool as before. We also mined the data more deeply using descriptive statistics (i.e., mean, median, standard deviation, minimum and maximum range of scores, skewness, and normality). This was an extra effort to present the data of the second condition in a more meaningful way and highlight potential relationships between them. Finally, concluding our quality assessment, we assessed the inter-rater agreement by checklist item, summing the common rating scores of raters' pairs for each item separately. For this, we divided them by the total number of quantitative or qualitative articles and multiplied them by 100 to produce the observed agreement as a percentage. Having observed the common checklist items with a low and high percentage of agreement among the three averaged, cross-verified groups, we then interpreted with caution the methodologies in articles that were relatively weaker (defined as "limited" by Lee et al.[139]).

RESULTS

Power of sample and content descriptives

Although Google Scholar does not offer an arithmetic overview of published articles (per topic/year), we found that between 1995 and 2022, the absolute number of articles published on PubMed—the greatly representative platform of authoritative publications—related to the specific research field (keywords: music and neuroplasticity) was n = 473 (100% of published articles since the first article publication), while n = 36 (7.5%) in 2019 (Figure 1). With a 17.5 absolute average and 3.5% of corresponding average percentage of articles published per year, we concluded that the overall number of articles we finally obtained and focused on, after searching both Google Scholar and PubMed, are of comparable validity to substantiate and justify the further elaboration and generalization of our results.

Regarding the contents of the 89 articles, under the "Research Site" category, 60 articles were lab-based, 18 mixed, and three were not lab-based (Figure 2). Under the "Tools Used" category (referring only to the quantitative profile), there were 63 neuroimaging and 19 non-neuroimaging articles using relevant tools. The neuroimaging articles included fMRI, DTI, EEG, magnetic resonance imaging (MRI), diffusion-weighted imaging, electrocorticogram, and near-infrared spectroscopy, whereas non-neuroimaging articles included drug administration, experimental tests, musical and cognitive tasks, musical therapeutic intervention methods, noise dosimeters, audiometers, headphones, vibrating gloves, neuropsychological and neurological tasks, behavioral tests for animals, and electrocardiograms (Figure 3). Finally, the "Participants Sample" range (number of participants) included six classes in total. Specifically, 15 articles were in the 1–20 number of participants range, 27 articles were in the 21–40 range, 20 articles were in the 41–60 range, five articles were in the 61–80 range, six articles were in the 81–100 range, and eight articles were in the 100+ participants range (Figure 4).

Reliability analysis

The result of Kendall's coefficient of concordance analysis was W = 0.44, $\chi^2 = 117.249$, $p = 0.020$, indicating a moderate agreement.[140] This result is in line with previously published IRR estimates obtained from coding similar constructs.[141] Based on these calculations, the null hypothesis was rejected, leading us to find that sufficient statistical evidence exists to infer that the three raters were concordant in their evaluations, and a modest amount of measurement error was introduced by them (as expected) due to differences in the subjective ratings given. The evaluations of the raters were comparable and were deemed suitable to help us answer the questions of the present study.

Continuing with the checklist items correlation matrix results, as calculated in our analysis protocol, these showed no strong correlation between the quantitative QualSyst checklist items for both the neuroimaging and the non-neuroimaging content profile groups (Tables S1 and S2). This is indicative of no causal relationship between the content of the checklist items. Furthermore, the correlation matrices showed that no checklist item had a significant impact on the raters' rating on another item. Based on these findings, we can confidently say that the raters' ratings have not been impacted by the contextual or serial order of the QualSyst checklist items.

Pairwise and multiple comparisons

Throughout the first condition analysis (pairwise comparisons: **Group 1**: raters A-B; **Group 2**: raters A-C; and **Group 3**: raters B-C), we observed some inter-rater agreement variations for the overall scores, the methodological quality, and by checklist items between the raters'

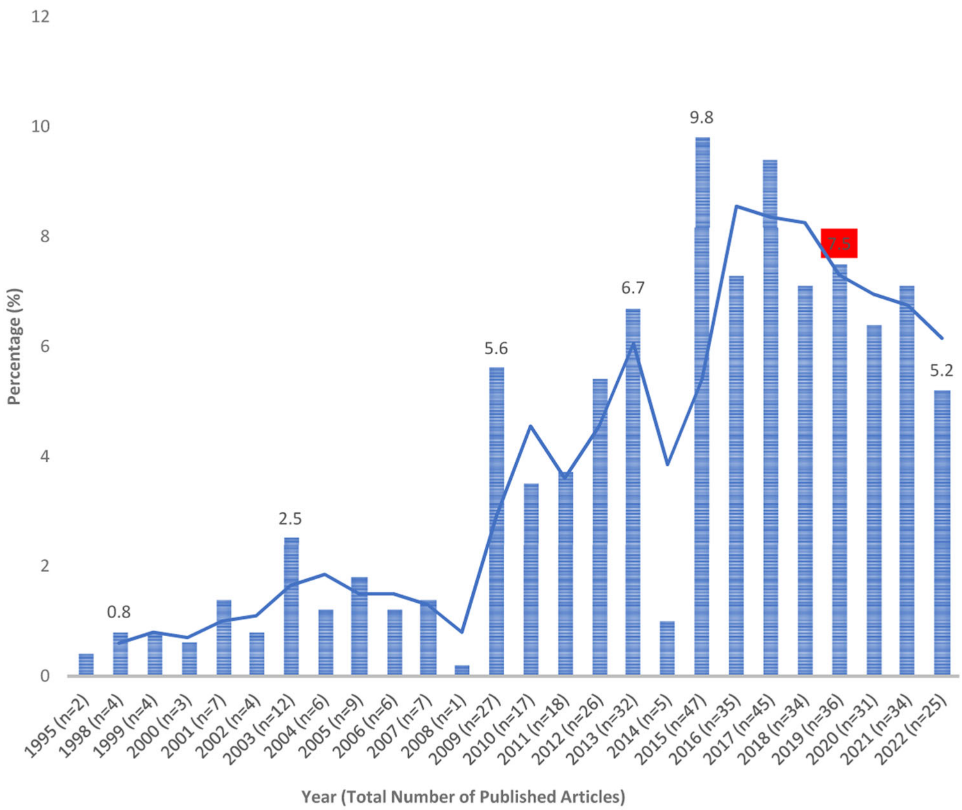

FIGURE 1 Total number of published articles according to PubMed between 1995 and 2022.

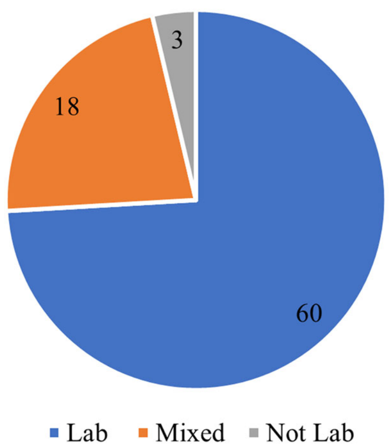

FIGURE 2 Number of articles based on Research Site content.

pairs. In brief, for the quantitative articles, the overall rating scores ranged from 0.23 to 1.0 (M = 0.675, SD = 0.12) for **Group 1**, from 0.23 to 1.0 (M = 0.725, SD = 0.11) for **Group 2,** and from 0.40 to 1.0 (M = 0.82, SD = 0.10) for **Group 3**. For the qualitative articles, the overall rating scores ranged from 0.2 to 1.0 (M = 0.61, SD = 0.23) for **Group 1**, from 0.2 to 1.0 (M = 0.625, SD = 0.19) for **Group 2**, and from 0.55 to 0.95 (M = 0.795, SD = 0.12) for **Group 3**. For an extensive presentation and critical analysis of the first condition results, see Appendix S1 and Table S3.

Although some relative variation was indeed observed in our first condition findings, our second more inclusive condition of analysis provided a more comprehensive methodological quality appraisal among all three raters' groups together (**Group 1** + **Group 2** + **Group 3**).

Inter-rater agreement for overall scores and methodological quality

Referring first to the quantitative articles (n = 81) separately, the histogram (Figure 5) and the boxplot (Figure 6) demonstrate that the distribution of the data is skewed to the left (skewness = −0.625, SE = 0.267) while being also more condensed. Thirty-eight articles (46.91%) were rated as good, 23 (28.39%) were rated as adequate, 19 (23.45%) were rated as strong, and only one (1.23%) was rated as limited. The mean quality score was 0.737 (SD = 0.084), with a minimum and maximum range from 0.44 to 0.96. The median was 0.75, indicating an overall good quality in these articles. Referring to the qualitative articles' quality scores, on the other hand, we observed (Figure 6) that the distribution of our data was quite symmetric (skewness = 0.148, SE = 0.752) and more spread out. Four papers (50%) were rated as adequate, two (25%) were rated as strong, one (12.5%) was rated as good, and only one (12.5%) was rated as limited. The mean quality score for these eight qualitative articles was 0.677 (SD = 0.144), with a range between 0.46 and 0.9. The median was 0.67, indicating an overall adequate quality. We should mention here, however, that the small sample of our data

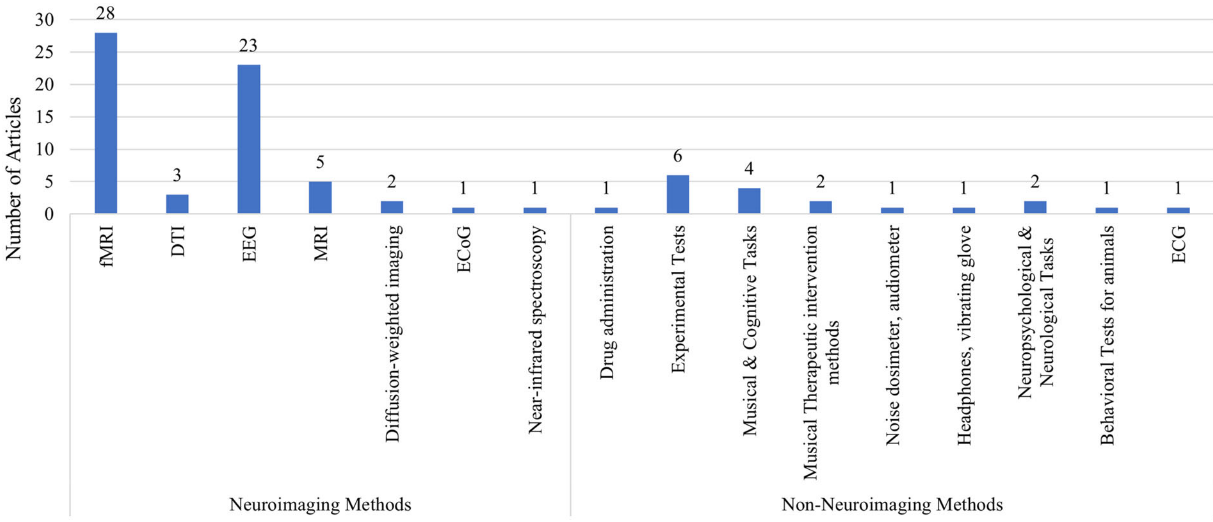

FIGURE 3 Number of articles based on Tools Used content.

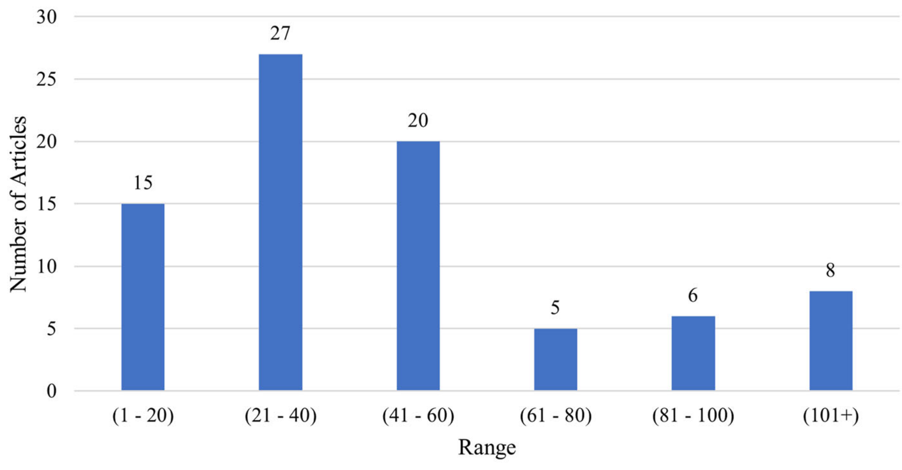

FIGURE 4 Number of articles based on Sample Interval content.

in this case ($n = 8$) prevented us from confidently evaluating normality, and, therefore, any relevant translation of results should be handled with caution.

Inter-rater agreement by checklist item

After examining the averaged scoring effect of separate checklist items for the quantitative articles, the research questions (checklist item 1), the study design (checklist item 2), the methods of subject selection (checklist item 3), the subject characteristics (checklist item 4), and the outcomes exposure measures (checklist item 8) seemed to be all sufficiently described and reported in all 81 articles. Also, almost all articles (79.8%) described and used appropriate analytic methods (checklist item 10). This confidently indicates that most analytic methods, such as chi-squared, *t*-tests, and the Kaplan–Meier with long rank tests (a common "survival analysis" technique used to analyze the time until an event of interest occurs), among others, were appropriately used in most published material. For example, Merten et al.[73] used chi-squared tests and univariate analyses of variance to explore possible relationships of the effect of motor synchronization on auditory discrimination performance experiment, and to what extent musical training and the brain anatomy of the auditory (Helsch's gyrus) and premotor areas affected both motor synchronization accuracy and melody discrimination. Also, in this same study, the researchers ran and adequately reported additional post-hoc analyses to examine individual differences in motor synchronization abilities and the brain anatomy of the frontal brain areas. On the other hand, a major scoring disagreement occurred between the raters on the topics of blinding (checklist item 7; 34.1%), sample size (checklist item 9; 37%), and confounders (checklist item 12; 24.6%).

More specifically, and according to the checklist-items rating guidelines of Kmet et al.,[136] the disagreement between the raters in "blinding" may have happened because of the four following reasons:

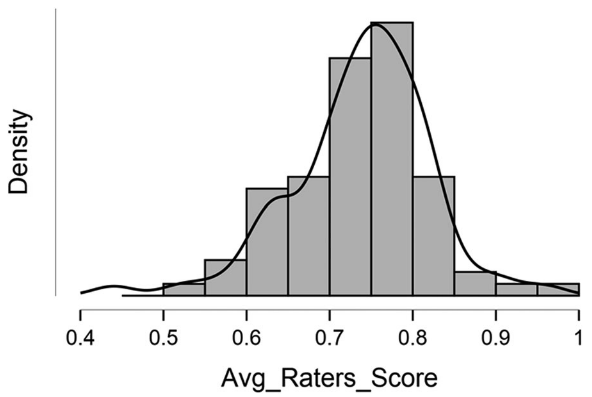

FIGURE 5 Distribution of scoring data for the quantitative articles.

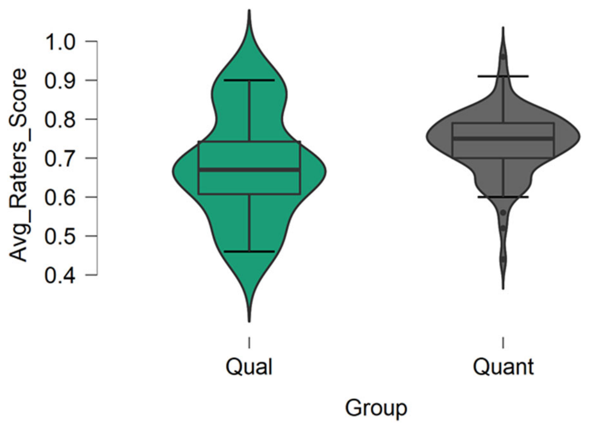

FIGURE 6 Distribution of scoring data for the quantitative and qualitative articles.

- Blinding was implicitly performed by the researchers but not explicitly reported (e.g., Ref. 119).
- Researchers conducted only observational studies (e.g., Refs. 64 and 127).
- Researchers performed experimental studies with no control groups included (e.g., Refs. 79 and 107).
- Researchers had only performed surveys in their studies (e.g., Ref. 113).

By the same token, the "sample size" (checklist item 9) was not deemed adequate for almost all quantitative articles, most probably due to the four following suggested reasons:

- Raters may have had insufficient data to clearly assess the sample size. The sample size may have seemed "small" and there was also no mention of the power effect size of interest, while variance estimates were not provided, too (e.g., Refs. 123 and 126).
- Technical considerations to rate high were imposed due to the presented small sample size arising from the participants' distinctive profile, which did not correlate with the study's selection criteria. For example, de Almeida et al.[100] evaluated the impact of music intervention during neonatal intensive care unit stay on 30 preterm and 15 full-term newborns' white matter maturation, using multi-modal MRI, as well as region-of-interest and seed-based tractography analyses. However, tract-specific analysis, an automated whole-brain diffusion-based group analysis technique, is generally considered appropriate for larger cohorts, and less suitable for infant populations due to the difficulty of computing robust medial tracts from the premature infant's thin sheet-like tracts.[142]
- There were articles in which the sample size seemed to be relatively "small," with statistically significant results and large effect sizes for all of these results (i.e., imprecise results) (e.g., Ref. 90).
- The sample size seemed to be obviously inadequate; meaning with some statistically nonsignificant results, without variance estimates, and most of the effect sizes large (e.g., Ref. 116).

Finally, most quantitative articles did not seem to be controlled for confounders (checklist item 12), either because research teams ran regression/mediation analyses without reporting any control for confounding variables (e.g., Ref. 91), or because they controlled for confounding variables but did not control for potential confounding factors (e.g., Refs. 95 and 120). Moreover, some researchers seemed to have presented only correlation (e.g., Ref. 109), test differences between groups (e.g., Refs. 111 and 114), and/or test differences between a single group, without, however, controlling for specific confounding variables (e.g., gender) that could have had an impact on their results (e.g., Ref. 72).

Regarding the qualitative articles, many of them gave a clear definition of the research question (checklist item 1) and the context (checklist item 3), while there is clear evidence of reflexivity (checklist item 10) in their reporting content. In addition, almost all qualitative articles (91.6%) adequately described and justified the theoretical framework informing the study and the methods used (checklist item 4). For example, Torppa and Huotilainen[79] explained in detail the reasons that music has a meaningful impact on speech and language skills enhancement for children with hearing impairments, based on previously published correlational and cross-sectional studies in this domain. Following the same reporting line, Papatzikis and Tsakmakidou[97] provided a detailed structural evaluation of their review, effectively enhancing the description of the impact of long-term effects of music exposure and early vocal contact on preterm infants' long-term neurodevelopment that Filippa et al.[143] debate in their book.

However, the analytic methods (checklist item 7; 25%) in the qualitative articles were not always clearly described, complete, and systematic. For example, although Leipold et al.[85] reported the neural mechanisms of perception and cognition in absolute pitch among musicians and nonmusicians based on specific previous published studies, it seems that they did not clearly describe their data analysis methods, leaving the reader wondering how the data were selected and processed.

DISCUSSION

As we have already presented in our methodology section, the electronic databases of Google Scholar and PubMed were thoroughly searched to identify all quantitative and qualitative studies published in 2019 for music and neuroplasticity research. After employing a strict set of inclusion and exclusion criteria, our aim was to explore the quality of these studies as reported by their authors in peer-reviewed publications. As a result of our work, we first and foremost realized that the interdisciplinary field of music and neuroplasticity is growing at an active pace and is hence worthwhile of being further supported at a more extensive and practical level. Furthermore, based on the diverse content we managed to obtain and analyze, our exploration is hoped to become an informative standard in specifying how music and neuroplasticity research should be further approached and carried out, following a well-defined, in terms of design and reporting, quality framework.

According to our findings, most of the studies in this specific research field were not only quantitative in nature, but a major proportion of it (46.91%) were deemed to be of good quality, too. When considering the quality threshold of 0.75 set by Lee et al.[139] for the accepted quality of quantitative articles in general literature, the average quality score of 0.737 for the current study's population certainly comes very close to it. If neuroplasticity is defined as a change that happens at a molecular, functional, or structural level, or in a skill such as processing speed or executive capacity, then these results not only suggest that we rightly follow this quantitative approach to assess this change, but that the quality of the reported research is of a good standard, too, upon which conclusions can safely be drawn regarding the association of music with brain plasticity.

Elaborating further on the quality of the analyzed quantitative articles, their research questions were found to be sufficiently well described, their study design evident and appropriate, their method of group selection appropriately described, the subject characteristics adequately specified, while the checklist item "analysis," after receiving the highest rating among the three raters' groups, confidently indicates appropriate use of analytic methods. However, it was also shown that the quantitative studies in the specific research field need to do better in controlling for confounders, do better in blinding or reporting the blinding of subjects, as well as to improve the sample size or at least the reporting of it. As an indication of the latter, the reported sample sizes for the quantitative studies included in our study were generally not large, with 50% of the articles having less than 40 participants—following the global norm for similar research[144]—and only less than 1% having over 100 participants.

As far as the qualitative articles are concerned, our study showed that these were less frequent in the literature. However, their methodological quality appeared to be sounder than their quantitative counterparts, scoring a rating of 25% "strong," with a mean quality score of 0.677, which is slightly higher than the global threshold of 0.55 set for qualitative articles by Lee et al.[139] This, in our opinion, suggests that the qualitative research conducted in the field seems to be sufficiently rooted in theoretical frameworks, being based on well-defined research questions, context, and reflexivity. Nevertheless, it is worth stressing that most of the qualitative articles did not meet the criteria of a very strong article. This indicates scope for methodological improvement in the design and practice of qualitative research methods by further improving the method of analysis and the reporting of the methods used in this research profile. In this way, we believe that more conclusive inferences can perhaps be drawn, by helping to discover, access, and further evaluate relevant results. Such a path may also mean better and further application of related quantitative data to various other research ecosystems, benefiting the advancement of developmental, educational, health, and medical practice, and the sciences as a whole.

RECOMMENDATIONS

Our study showed that researchers studying music and neuroplasticity generally follow good strategies for designing and reporting their research. However, the results also suggested that researchers should always take a closer look at the preparation, implementation, and reporting of certain aspects of their protocol, mainly the control of confounding factors, the sample size, as well as the blinding procedures they use to conduct their study. Any omissions in these areas, whether at the design or reporting level, can ultimately lead to questionable, nonrepeatable, and of course impractical results. In light of these, the following recommendations should be taken as much as possible into account to improve practices related to research endeavors in the specific field.

In terms of avoiding and controlling for confounding factors, research has generally shown that these can be achieved either during the design or during the analysis stage of a research study.[145] Regarding the design stage, especially when working on brain and neuroplasticity, the golden rule is to follow a randomized control trial (RCT) protocol and then clearly report it, too.[146] Following this RCT approach, we essentially create experimental groups that consist of—as much as possible—the same participant profiles, differing only in the variables studied. On the other hand, controlling for and avoiding the incorporation of confounders during the analysis phase of a study can be seen as a backup plan to the previous step. This approach has been found to involve either the more modern multivariate coding analyses or the traditional mass-univariate analyses.[147] In whichever case that might be, great care is needed in their application, as strong biases can be easily introduced due to, for example, subjective input on population-level parameters.

As far as the improvement of the participant sample size and its reporting is concerned, two indirect but effective recommendations have been shown to work very well so far in neuroscientific research, particularly in the field of neuroimaging. These are the preregistration of studies in scientific journals, as well as the (pre)calculation and establishment of the sample size. The calculation of sample size can either happen through the pre-estimation of the sample power at a technical level, or at a policy level by explicitly establishing incentives provided by specific funding bodies aiming at targeted sample numbers for certain reasons. A meta-analysis by Szucs and Ioannidis[144] on the evolution of participant sample sizes in neuroimaging research from 1990 to 2018 suggested, on the one hand, that precalculation of the sample power seems to be a very effective strategy to eventually achieve realistic effect sizes during the data collection and analysis stages of a study. They also suggested that the preregistration of a study can help avoid rephrasing of protocol features, and thus changing sample recruitment strategies at will without a previous clear plan, reason, and explanation.

Finally, with regard to the blinding procedures, according to the literature, the rule of thumb recommends that the more people (i.e., participants, researchers, and technical staff) who can be blinded to the content, research techniques, and the data collection and analysis procedures, the better it is for the study's final result. According to previous work by Mataix-Cols and Andersson[148] and Basoglu et al.,[149] researchers should try to reveal to participants as little information as possible about the research protocol they are involved in, but always within the boundaries of what is ethically acceptable and correct; to ensure in any way possible that participants do not reveal their group allocation to their assessors when and if they come into contact; to recruit independent external assessors, blinded to the research hypothesis and the interventions used in the trials; to ensure that data managers and statisticians are blinded to group allocations; to apply blinded interpretation of data to reduce interpretive bias (e.g., by creating two abstracts of data interpretation that are blinded to the team, choosing the correct approach after the breaking of blinding); and to report these procedures with as much information as possible in the final report for further clarifications or study duplication.

CONCLUSIONS

We believe that the current paper can set an example for examining the quality of studies in the area of music and neuroplasticity when using a rigorous system of evaluation. The multiple levels of reliability checking of the rater agreement and the cross-referencing of scoring as presented here stand up to similar tests devised by others. Thus, any relevant approach of evaluation may not only be reliable to raise evidence for systematic reviews or meta-analyses in the specific field, but also to evaluate the quality of the articles composing this field. Also, this extended evaluation approach may not only be valuable at an individual researcher level—to know what to expect, use, and apply in terms of future research endeavors—but also at a more collective level, where specific research directions are discussed and needed without the cost of undefined qualitative research standards. If nothing else, it seems that at a specific domain level, the influence of music on neuroplasticity is justified by the quality of the research and the researchers' knowledge on which it is based. Therefore, we believe[150] that more work like this one is further needed to present new technically critical content and processes that will certainly help to further improve the quality of the field's research outcomes.

AUTHOR CONTRIBUTIONS

E.P. conceptualized and supervised the work, developed the analysis pipeline, and analyzed the data with input from M.A., R.N.S., V.P., and F.Z. M.A., R.N.S., and V.P. visualized the results. E.P. wrote the manuscript with help from M.A., R.N.S., V.P., and F.Z.

COMPETING INTERESTS

The authors declare no competing interests.

ORCID

Efthymios Papatzikis https://orcid.org/0000-0003-1801-6986

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/nyas.14944.

REFERENCES

1. Sasmita, A. O., Kuruvilla, J., & Ling, A. P. (2018). Harnessing neuroplasticity: Modern approaches and clinical future. *International Journal of Neuroscience*, *128*(11), 1061–1077.
2. Merrett, D. L., Peretz, I., & Wilson, S. J. (2013). Moderating variables of music training-induced neuroplasticity: A review and discussion. *Frontiers in Psychology*, *4*, 606.
3. Stegemöller, E. L. (2014). Exploring a neuroplasticity model of music therapy. *Journal of Music Therapy*, *51*(3), 211–227.
4. Fuchs, E., & Flügge, G. (2014). Adult neuroplasticity: More than 40 years of research. *Neural Plasticity*, 541870.
5. Kraus, C., Castrén, E., Kasper, S., & Lanzenberger, R. (2017). Serotonin and neuroplasticity–links between molecular, functional and structural pathophysiology in depression. *Neuroscience & Biobehavioral Reviews*, *77*, 317–326.
6. Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48.
7. Reybrouck, M., & Eerola, T. (2017). Music and its inductive power: A psychobiological and evolutionary approach to musical emotions. *Frontiers in Psychology*, *8*, 494.
8. Reybrouck, M., & Brattico, E. (2015). Neuroplasticity beyond sounds: Neural adaptations following long-term musical aesthetic experiences. *Brain Sciences*, *5*(1), 69–91.
9. Trehub, S. E. (2003). Toward a developmental psychology of music. *Annals of the New York Academy of Sciences*, *999*(1), 402–413.
10. Miller, G. (2000). Evolution of human music through sexual selection. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 329–360). MIT Press.
11. Papatzikis, E., & Rishony, H. (2022). What is music for neuroplasticity?: Combined value on infant development and inclusion. In M. Efstratopoulou (Ed.) *Rethinking inclusion and transformation in special education* (pp. 160–177). New York: IGI Global.
12. Chorna, O., Filippa, M., De Almeida, J. S., Lordier, L., Monaci, M. G., Hüppi, P., Grandjean, D., & Guzzetta, A. (2019). Neuroprocessing mechanisms of music during fetal and neonatal development: A role in neuroplasticity and neurodevelopment. *Neural Plasticity*, *2019*, 3972918.
13. Medina, D., & Barraza, P. (2019). Efficiency of attentional networks in musicians and non-musicians. *Heliyon*, *5*(3), e01315.
14. Acer, N., Bastepe-Gray, S., Sagiroglu, A., Gumus, K. Z., Degirmencioglu, L., Zararsiz, G., & Ozic, M. U. (2018). Diffusion tensor and volumetric magnetic resonance imaging findings in the brains of professional musicians. *Journal of Chemical Neuroanatomy*, *88*, 33–40.

15. Burunat, I., Brattico, E., Puoliväli, T., Ristaniemi, T., Sams, M., & Toiviainen, P. (2015). Action in perception: Prominent visuo-motor functional symmetry in musicians during music listening. *PLoS ONE*, *10*(9), e0138238.
16. Lee, D. J., Chen, Y., & Schlaug, G. (2003). Corpus callosum: Musician and gender effects. *Neuroreport*, *14*(2), 205–209.
17. Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*(8), 1047–1055.
18. Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*, *117*(4), 1924–1934.
19. Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59.
20. Madison, G., Holmquist, J., & Vestin, M. (2018). Musical improvisation skill in a prospective partner is associated with mate value and preferences, consistent with sexual selection and parental investment theory: Implications for the origin of music. *Evolution and Human Behavior*, *39*(1), 120–129.
21. Snowdon, C. T., Zimmermann, E., & Altenmüller, E. (2015). Music evolution and neuroscience. *Progress in Brain Research*, *217*, 17–34.
22. Carter, T. E., & Panisch, L. S. (2021). A systematic review of music therapy for psychosocial outcomes of substance use clients. *International Journal of Mental Health and Addiction*, *19*, 1551–1568.
23. Mofredj, A., Alaya, S., Tassaioust, K., Bahloul, H., & Mrabet, A. (2016). Music therapy, a review of the potential therapeutic benefits for the critically ill. *Journal of Critical Care*, *35*, 195–199.
24. Kamioka, H., Tsutani, K., Yamada, M., Park, H., Okuizumi, H., Tsuruoka, K., Honda, T., Okada, S., Park, S. J., Kitayuguchi, J., & Abe, T. (2014). Effectiveness of music therapy: A summary of systematic reviews based on randomized controlled trials of music interventions. *Patient Preference and Adherence*, *8*, 727.
25. Miendlarzewska, E. A., & Trost, W. J. (2014). How musical training affects cognitive development: Rhythm, reward and other modulating variables. *Frontiers in Neuroscience*, *7*, 279.
26. Lordier, L., Meskaldji, D. E., Grouiller, F., Pittet, M. P., Vollenweider, A., Vasung, L., Borradori-Tolsa, C., Lazeyras, F., Grandjean, D., Van De Ville, D., & Hüppi, P. S. (2019). Music in premature infants enhances high-level cognitive brain networks. *Proceedings of the National Academy of Sciences*, *116*(24), 12103–12108.
27. Papatzikis, E. (2021). An infant's question on COVID-19 and music: Should I attend my online classes? *Frontiers in Psychology*, *12*, 771050.
28. Swaminathan, S., & Schellenberg, E. G. (2018). Music training and cognitive abilities: Associations, causes, and consequences. In M. Thaut, & D. A. Hodges (Eds.) *The Oxford handbook of music and neuroscience*. 645–670.
29. Costa-Giomi, E. (2015). The long-term effects of childhood music instruction on intelligence and general cognitive abilities. *Applications of Research in Music Education*, *33*(2), 20–26.
30. Schellenberg, E. G. (2020). Music training, individual differences, and plasticity. In M. S. C. Thomas, D. Mareschal, & I. Dumontheil (Eds.) *Educational neuroscience* (pp. 415–441). Routledge.
31. Kuhn, T. (2012). *The structure of scientific revolutions; With an introductory essay by Ian Hacking*. University of Chicago Press.
32. Callard, F., & Fitzgerald, D. (2015). *Rethinking interdisciplinarity across the social sciences and neurosciences*. Palgrave Macmillan.
33. Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R., & Chen, J. H. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, *57*(7), 1798–1806.
34. Cross, K., & Fujioka, T. (2019). Auditory rhyme processing in expert freestyle rap lyricists and novices: An ERP study. *Neuropsychologia*, *129*, 223–235.
35. Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P., & Limb, C. J. (2014). Neural substrates of interactive musical improvisation: An FMRI study of 'trading fours' in jazz. *PLoS ONE*, *9*(2), e88665.
36. Moore, E., Schaefer, R. S., Bastin, M. E., Roberts, N., & Overy, K. (2014). Can musical training influence brain connectivity? Evidence from diffusion tensor MRI. *Brain Sciences*, *4*(2), 405–427.
37. Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage*, *17*(3), 1613–1622.
38. Chacon-Castano, J., Rathbone, D. R., Hoffman, R., Yang, H., Pantazis, D., Yang, J., Hornberger, E., & Hanumara, N. C. (2017). Music and the brain-design of an MEG compatible piano. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society *(EMBC)* (pp. 521–524). IEEE.
39. Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, *23*(27), 9240–9245.
40. Vieillard, S., & Gilet, A. L. (2013). Age-related differences in affective responses to and memory for emotions conveyed by music: A cross-sectional study. *Frontiers in Psychology*, *4*, 711.
41. Jaschke, A. C., Eggermont, L. H., Honing, H., & Scherder, E. J. (2013). Music education and its effect on intellectual abilities in children: A systematic review. *Reviews in the Neurosciences*, *24*(6), 665–675.
42. Fang, R., Ye, S., Huangfu, J., & Calimag, D. P. (2017). Music therapy is a potential intervention for cognition of Alzheimer's disease: A mini-review. *Translational Neurodegeneration*, *6*(1), 1–8.
43. Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, *48*(8), 1429–1441.
44. Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186.
45. Chacón-Moscoso, S., Sanduvete-Chaves, S., & Sánchez-Martín, M. (2016). The development of a checklist to enhance methodological quality in intervention programs. *Frontiers in Psychology*, *7*, 1811.
46. Buring, J. E. (1987). *Epidemiology in medicine*. Lippincott Williams & Wilkins.
47. Shany, O., Singer, N., Gold, B. P., Jacoby, N., Tarrasch, R., Hendler, T., & Granot, R. (2019). Surprise-related activation in the nucleus accumbens interacts with music-induced pleasantness. *Social Cognitive and Affective Neuroscience*, *14*(4), 459–470.
48. Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J., & Marco-Pallarés, J. (2019). White matter microstructure reflects individual differences in music reward sensitivity. *Journal of Neuroscience*, *39*(25), 5018–5027.
49. Cheung, V. K., Harrison, P. M., Meyer, L., Pearce, M. T., Haynes, J. D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology*, *29*(23), 4084–4092.
50. Gold, B. P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A., & Zatorre, R. J. (2019). Musical reward prediction errors engage the nucleus accumbens and motivate learning. *Proceedings of the National Academy of Sciences*, *116*(8), 3310–3315.
51. Belfi, A. M., & Loui, P. (2020). Musical anhedonia and rewards of music listening: Current advances and a proposed model. *Annals of the New York Academy of Sciences*, *1464*(1), 99–114.
52. Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., Olivé, G., Marco-Pallarés, J., Antonijoan, R. M., Valle, M., & Riba, J. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences*, *116*(9), 3793–3798.

53. Kim, S. G., Mueller, K., Lepsien, J., Mildner, T., & Fritz, T. H. (2019). Brain networks underlying aesthetic appreciation as modulated by interaction of the spectral and temporal organisations of music. *Scientific Reports*, *9*(1), 1–5.
54. Nemati, S., Akrami, H., Salehi, S., Esteky, H., & Moghimi, S. (2019). Lost in music: Neural signature of pleasure and its role in modulating attentional resources. *Brain Research*, *1711*, 7–15.
55. Borges, A. F., Irrmischer, M., Brockmeier, T., Smit, D. J., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2019). Scaling behaviour in music and cortical dynamics interplay to mediate music listening pleasure. *Scientific Reports*, *9*(1), 1–5.
56. Klepzig, K., Horn, U., König, J., Holtz, K., Wendt, J., Hamm, A. O., & Lotze, M. (2020). Brain imaging of chill reactions to pleasant and unpleasant sounds. *Behavioural Brain Research*, *380*, 112417.
57. Esfahani-Bayerl, N., Finke, C., Kopp, U., Moon, D. U., & Ploner, C. J. (2019). Musical memory and hippocampus revisited: Evidence from a musical layperson with highly selective hippocampal damage. *Cortex*, *119*, 519–527.
58. Karmonik, C., Brandt, A., Elias, S., Townsend, J., Silverman, E., Shi, Z., & Frazier, J. T. (2020). Similarity of individual functional brain connectivity patterns formed by music listening quantified with a data-driven approach. *International Journal of Computer Assisted Radiology and Surgery*, *15*(4), 703–713.
59. Li, Q., Wang, X., Wang, S., Xie, Y., Li, X., Xie, Y., & Li, S. (2019). Dynamic reconfiguration of the functional brain network after musical training in young adults. *Brain Structure and Function*, *224*(5), 1781–1795.
60. Belden, A., Zeng, T., Przysinda, E., Anteraper, S. A., Whitfield-Gabrieli, S., & Loui, P. (2020). Improvising at rest: Differentiating jazz and classical music training with resting state functional connectivity. *Neuroimage*, *207*, 116384.
61. Dhakal, K., Norgaard, M., Adhikari, B. M., Yun, K. S., & Dhamala, M. (2019). Higher node activity with less functional connectivity during musical improvisation. *Brain Connectivity*, *9*(3), 296–309.
62. Gujing, L., Hui, H., Xin, L., Lirong, Z., Yutong, Y., Guofeng, Y., Jing, L., Shulin, Z., Lei, Y., Cheng, L., & Dezhong, Y. (2019). Increased insular connectivity and enhanced empathic ability associated with dance/music training. *Neural Plasticity*, *2019*, 9693109.
63. Zamorano, A. M., Montoya, P., Cifre, I., Vuust, P., Riquelme, I., & Kleber, B. (2019). Experience-dependent neuroplasticity in trained musicians modulates the effects of chronic pain on insula-based networks–A resting-state fMRI study. *Neuroimage*, *202*, 116103.
64. Lumaca, M., Kleber, B., Brattico, E., Vuust, P., & Baggio, G. (2019). Functional connectivity in human auditory networks and the origins of variation in the transmission of musical systems. *eLife*, *8*, e48710.
65. Tseng, Y. L., Liu, H. H., Liou, M., Tsai, A. C., Chien, V. S., Shyu, S. T., & Yang, Z. S. (2019). Lingering sound: Event-related phase-amplitude coupling and phase-locking in fronto-temporo-parietal functional networks during memory retrieval of music melodies. *Frontiers in Human Neuroscience*, *13*, 150.
66. Wu, K., Anderson, J., Townsend, J., Frazier, T., Brandt, A., & Karmonik, C. (2019). Characterization of functional brain connectivity towards optimization of music selection for therapy: A fMRI study. *International Journal of Neuroscience*, *129*(9), 882–889.
67. Tanaka, S., & Kirino, E. (2019). Increased functional connectivity of the angular gyrus during imagined music performance. *Frontiers in Human Neuroscience*, *13*, 92.
68. Bidelman, G. M., & Walker, B. (2019). Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. *Neuroimage*, *201*, 116022.
69. Daly, I., Williams, D., Hwang, F., Kirke, A., Miranda, E. R., & Nasuto, S. J. (2019). Electroencephalography reflects the activity of sub-cortical brain regions during approach-withdrawal behaviour while listening to music. *Scientific Reports*, *9*(1), 1–22.
70. Aydın, S., Güdücü, Ç., Kutluk, F., Öniz, A., & Özgören, M. (2019). The impact of musical experience on neural sound encoding performance. *Neuroscience Letters*, *694*, 124–128.
71. Wang, W., Wei, L., Chen, N., Jones, J. A., Gong, G., & Liu, H. (2019). Decreased gray-matter volume in insular cortex as a correlate of singers' enhanced sensorimotor control of vocal production. *Frontiers in Neuroscience*, *13*, 815.
72. Celma-Miralles, A., & Toro, J. M. (2019). Ternary meter from spatial sounds: Differences in neural entrainment between musicians and non-musicians. *Brain and Cognition*, *136*, 103594.
73. Merten, N., Kramme, J., Breteler, M., & Herholz, S. C. (2019). Previous musical experience and cortical thickness relate to the beneficial effect of motor synchronization on auditory function. *Frontiers in Neuroscience*, *13*, 1042.
74. Rajan, A., Valla, J. M., Alappatt, J. A., Sharda, M., Shah, A., Ingalhalikar, M., & Singh, N. C. (2019). Wired for musical rhythm? A diffusion MRI-based study of individual differences in music perception. *Brain Structure and Function*, *224*(5), 1711–1722.
75. Mankel, K., Barber, J., & Bidelman, G. M. (2020). Auditory categorical processing for speech is modulated by inherent musical listening skills. *Neuroreport*, *31*(2), 162.
76. Angulo-Perkins, A., & Concha, L. (2019). Discerning the functional networks behind processing of music and speech through human vocalizations. *PLoS ONE*, *14*(10), e0222796.
77. Torppa, R., Faulkner, A., Laasonen, M., Lipsanen, J., & Sammler, D. (2020). Links of prosodic stress perception and musical activities to language skills of children with cochlear implants and normal hearing. *Ear and Hearing*, *41*(2), 395–410.
78. Tillmann, B., Poulin-Charronnat, B., Gaudrain, E., Akhoun, I., Delbé, C., Truy, E., & Collet, L. (2019). Implicit processing of pitch in postlingually deafened cochlear implant users. *Frontiers in Psychology*, *10*, 1990.
79. Torppa, R., & Huotilainen, M. (2019). Why and how music can be used to rehabilitate and develop speech and language skills in hearing-impaired children. *Hearing Research*, *380*, 108–122.
80. Leipold, S., Brauchli, C., Greber, M., & Jäncke, L. (2019). Absolute and relative pitch processing in the human brain: Neural and behavioral evidence. *Brain Structure and Function*, *224*(5), 1723–1738.
81. Brauchli, C., Leipold, S., & Jäncke, L. (2019). Univariate and multivariate analyses of functional networks in absolute pitch. *Neuroimage*, *189*, 241–247.
82. Sharma, V. V., Thaut, M., Russo, F., & Alain, C. (2019). Absolute pitch and musical expertise modulate neuro-electric and behavioral responses in an auditory Stroop paradigm. *Frontiers in Neuroscience*, *13*, 932.
83. Burkhard, A., Elmer, S., & Jäncke, L. (2019). Early tone categorization in absolute pitch musicians is subserved by the right-sided perisylvian brain. *Scientific Reports*, *9*(1), 1–4.
84. McKetton, L., DeSimone, K., & Schneider, K. A. (2019). Larger auditory cortical area and broader frequency tuning underlie absolute pitch. *Journal of Neuroscience*, *39*(15), 2930–2937.
85. Leipold, S., Greber, M., & Elmer, S. (2019). Perception and cognition in absolute pitch: Distinct yet inseparable. *Journal of Neuroscience*, *39*(30), 5839–5841.
86. Matsuda, M., Igarashi, H., & Itoh, K. (2019). Auditory T-complex reveals reduced neural activities in the right auditory cortex in musicians with absolute pitch. *Frontiers in Neuroscience*, *13*, 809.
87. Wenhart, T., Bethlehem, R. A., Baron-Cohen, S., & Altenmueller, E. (2019). Autistic traits, resting-state connectivity, and absolute pitch in professional musicians: Shared and distinct neural features. *Molecular Autism*, *10*(1), 1–8.
88. Leipold, S., Greber, M., Sele, S., & Jäncke, L. (2019). Neural patterns reveal single-trial information on absolute pitch and relative pitch perception. *Neuroimage*, *200*, 132–141.

89. Coll, S. Y., Vuichoud, N., Grandjean, D., & James, C. E. (2019). Electrical neuroimaging of music processing in pianists with and without true absolute pitch. *Frontiers in Neuroscience*, *13*, 142.
90. Skoe, E., Camera, S., & Tufts, J. (2019). Noise exposure may diminish the musician advantage for perceiving speech in noise. *Ear and Hearing*, *40*(4), 782–793.
91. Zendel, B. R., West, G. L., Belleville, S., & Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiology of Aging*, *81*, 102–115.
92. Fleming, D., Belleville, S., Peretz, I., West, G., & Zendel, B. R. (2019). The effects of short-term musical training on the neural processing of speech-in-noise in older adults. *Brain and Cognition*, *136*, 103592.
93. Hernández, M., Palomar-García, M. Á., Nohales-Nieto, B., Olcina-Sempere, G., Villar-Rodríguez, E., Pastor, R., Ávila, C., & Parcet, M. A. (2019). Separate contribution of striatum volume and pitch discrimination to individual differences in music reward. *Psychological Science*, *30*(9), 1352–1361.
94. Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology*, *36*(1–2), 1–7.
95. Sharp, A., Houde, M. S., Maheu, M., Ibrahim, I., & Champoux, F. (2019). Improved tactile frequency discrimination in musicians. *Experimental Brain Research*, *237*(6), 1575–1580.
96. Filippa, M., Lordier, L., De Almeida, J. S., Monaci, M. G., Adam-Darque, A., Grandjean, D., Kuhn, P., & Hüppi, P. S. (2020). Early vocal contact and music in the NICU: New insights into preventive interventions. *Pediatric Research*, *87*(2), 249–264.
97. Papatzikis, E., & Tsakmakidou, N. (2021). Book review: Manuela Filippa, Pierre Kuhn, & Björn Westrup. Early vocal contact and preterm infant brain development: Bridging the gaps between research and practice. *Psychology of Music*, *49*(2), 304–308.
98. Lordier, L., Loukas, S., Grouiller, F., Vollenweider, A., Vasung, L., Meskaldij, D. E., Lejeune, F., Pittet, M. P., Borradori-Tolsa, C., Lazeyras, F., & Grandjean, D. (2019). Music processing in preterm and full-term newborns: A psychophysiological interaction (PPI) approach in neonatal fMRI. *Neuroimage*, *185*, 857–864.
99. Lejeune, F., Lordier, L., Pittet, M. P., Schoenhals, L., Grandjean, D., Hüppi, P. S., Filippa, M., & Borradori Tolsa, C. (2019). Effects of an early postnatal music intervention on cognitive and emotional development in preterm children at 12 and 24 months: Preliminary findings. *Frontiers in Psychology*, *10*, 494.
100. de Almeida, J. S., Lordier, L., Zollinger, B., Kunz, N., Bastiani, M., Gui, L., Adam-Darque, A., Borradori-Tolsa, C., Lazeyras, F., & Hüppi, P. S. (2020). Music enhances structural maturation of emotional processing neural pathways in very preterm infants. *Neuroimage*, *207*, 116391.
101. Suppanen, E., Huotilainen, M., & Ylinen, S. (2019). Rhythmic structure facilitates learning from auditory input in newborn infants. *Infant Behavior and Development*, *57*, 01346.
102. Loui, P., Raine, L. B., Chaddock-Heyman, L., Kramer, A. F., & Hillman, C. H. (2019). Musical instrument practice predicts white matter microstructure and cognitive abilities in childhood. *Frontiers in Psychology*, *10*, 1198.
103. Hennessy, S. L., Sachs, M. E., Ilari, B., & Habibi, A. (2019). Effects of music training on inhibitory control and associated neural networks in school-aged children: A longitudinal study. *Frontiers in Neuroscience*, *13*, 1080.
104. Putkinen, V., Huotilainen, M., & Tervaniemi, M. (2019). Neural encoding of pitch direction is enhanced in musically trained children and is related to reading skills. *Frontiers in Psychology*, *10*, 1475.
105. Putkinen, V., Tervaniemi, M., & Huotilainen, M. (2019). Musical playschool activities are linked to faster auditory development during preschool-age: A longitudinal ERP study. *Scientific Reports*, *9*(1), 11310.
106. Politimou, N., Dalla Bella, S., Farrugia, N., & Franco, F. (2019). Born to speak and sing: Musical predictors of language development in preschoolers. *Frontiers in Psychology*, *10*, 948.
107. Barbaroux, M., Dittinger, E., & Besson, M. (2019). Music training with Démos program positively influences cognitive functions in children from low socio-economic backgrounds. *PLoS ONE*, *14*(5), e0216874.
108. Fasano, M. C., Semeraro, C., Cassibba, R., Kringelbach, M. L., Monacis, L., de Palo, V., Vuust, P., & Brattico, E. (2019). Short-term orchestral music training modulates hyperactivity and inhibitory control in school-age children: A longitudinal behavioural study. *Frontiers in Psychology*, *10*, 750.
109. Ireland, K., Iyer, T. A., & Penhune, V. B. (2019). Contributions of age of start, cognitive abilities and practice to musical task performance in childhood. *PLoS ONE*, *14*(4), e0216119.
110. Strong, J. V., & Mast, B. T. (2019). The cognitive functioning of older adult instrumental musicians and non-musicians. *Aging, Neuropsychology, and Cognition*, *26*(3), 367–386.
111. Saarikivi, K. A., Huotilainen, M., Tervaniemi, M., & Putkinen, V. (2019). Selectively enhanced development of working memory in musically trained children and adolescents. *Frontiers in Integrative Neuroscience*, *13*, 62.
112. Bird, L. J., Jackson, G. D., & Wilson, S. J. (2019). Music training is neuroprotective for verbal cognition in focal epilepsy. *Brain*, *142*(7), 1973–1987.
113. LaGasse, A. B., Manning, R. C., Crasta, J. E., Gavin, W. J., & Davies, P. L. (2019). Assessing the impact of music therapy on sensory gating and attention in children with autism: A pilot and feasibility study. *Journal of Music Therapy*, *56*(3), 287–314.
114. Siponkoski, S. T., Martínez-Molina, N., Kuusela, L., Laitinen, S., Holma, M., Ahlfors, M., Jordan-Kilkki, P., Ala-Kauhaluoma, K., Melkas, S., Pekkola, J., & Rodriguez-Fornells, A. (2020). Music therapy enhances executive functions and prefrontal structural neuroplasticity after traumatic brain injury: Evidence from a randomized controlled trial. *Journal of Neurotrauma*, *37*(4), 618–634.
115. Vik, B. M. (2019). Music-supported systematic treatment strategy for patients with executive dysfunction following traumatic brain injury: Similarities and divergencies in 7 case reports. *Music and Medicine*, *11*(3), 166–175.
116. Vik, B. M., Skeie, G. O., & Specht, K. (2019). Neuroplastic effects in patients with traumatic brain injury after music-supported therapy. *Frontiers in Human Neuroscience*, *13*, 177.
117. Puyjarinet, F., Bégel, V., Gény, C., Driss, V., Cuartero, M. C., Kotz, S. A., Pinto, S., & Bella, D. S. (2019). Heightened orofacial, manual, and gait variability in Parkinson's disease results from a general rhythmic impairment. *Npj Parkinson's Disease*, *5*(1), 1–7.
118. Tamplin, J., Morris, M. E., Marigliani, C., Baker, F. A., & Vogel, A. P. (2019). ParkinSong: A controlled trial of singing-based therapy for Parkinson's disease. *Neurorehabilitation and Neural Repair*, *33*(6), 453–463.
119. Slattery, C. F., Agustus, J. L., Paterson, R. W., McCallion, O., Foulkes, A. J., Macpherson, K., Carton, A. M., Harding, E., Golden, H. L., Jaisin, K., & Mummery, C. J. (2019). The functional neuroanatomy of musical memory in Alzheimer's disease. *Cortex*, *115*, 357–370.
120. King, J. B., Jones, K. G., Goldberg, E., Rollins, M., MacNamee, K., Moffit, C., Naidu, S. R., Ferguson, M. A., Garcia-Leavitt, E., Amaro, J., & Breitenbach, K. R. (2019). Increased functional connectivity after listening to favored music in adults with Alzheimer dementia. *Journal of Prevention of Alzheimer's Disease*, *6*(1), 56–62.
121. Leggieri, M., Thaut, M. H., Fornazzari, L., Schweizer, T. A., Barfett, J., Munoz, D. G., & Fischer, C. E. (2019). Music intervention approaches for Alzheimer's disease: A review of the literature. *Frontiers in Neuroscience*, *13*, 132.
122. Mathias, B., Gehring, W. J., & Palmer, C. (2019). Electrical brain responses reveal sequential constraints on planning during music performance. *Brain Sciences*, *9*(2), 25.

123. Zhu, Y., Liu, J., Mathiak, K., Ristaniemi, T., & Cong, F. (2019). Deriving electrophysiological brain network connectivity via tensor component analysis during freely listening to music. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(2), 409–418.
124. Ding, Y., Zhang, Y., Zhou, W., Ling, Z., Huang, J., Hong, B., & Wang, X. (2019). Neural correlates of music listening and recall in the human brain. *Journal of Neuroscience*, *39*(41), 8112–8123.
125. Alain, C., Moussard, A., Singer, J., Lee, Y., Bidelman, G. M., & Moreno, S. (2019). Music and visual art training modulate brain activity in older adults. *Frontiers in Neuroscience*, *13*, 182.
126. Feng, K., Shen, C. Y., Ma, X. Y., Chen, G. F., Zhang, M. L., Xu, B., Liu, X. M., Sun, J. J., Zhang, X. Q., Liu, P. Z., & Ju, Y. (2019). Effects of music therapy on major depressive disorder: A study of prefrontal hemodynamic functions using fNIRS. *Psychiatry Research*, *275*, 86–93.
127. Giacosa, C., Karpati, F. J., Foster, N. E., Hyde, K. L., & Penhune, V. B. (2019). The descending motor tracts are different in dancers and musicians. *Brain Structure and Function*, *224*(9), 3229–3246.
128. Harding, E. E., Sammler, D., Henry, M. J., Large, E. W., & Kotz, S. A. (2019). Cortical tracking of rhythm in music and speech. *Neuroimage*, *185*, 96–101.
129. Jiam, N. T., Deroche, M. L., Jiradejvong, P., & Limb, C. J. (2019). A randomized controlled crossover study of the impact of online music training on pitch and timbre perception in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, *20*(3), 247–262.
130. Mencke, I., Omigie, D., Wald-Fuhrmann, M., & Brattico, E. (2019). Atonal music: Can uncertainty lead to pleasure? *Frontiers in Neuroscience*, *12*, 979.
131. Norman-Haignere, S. V., Kanwisher, N., McDermott, J. H., & Conway, B. R. (2019). Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. *Nature Neuroscience*, *22*(7), 1057–1060.
132. Papadakakis, A., Sidiropoulou, K., & Panagis, G. (2019). Music exposure attenuates anxiety-and depression-like behaviors and increases hippocampal spine density in male rats. *Behavioural Brain Research*, *372*, 112023.
133. Kmet, L. M., Cook, L. S., & Lee, R. C. (2004). *Standard quality assessment criteria for evaluating primary research papers from a variety of fields.* Alberta Heritage Foundation for Medical Research (AHFMR).
134. Cho, M. K., & Bero, L. A. (1994). Instruments for assessing the quality of drug studies published in the medical literature. *JAMA*, *272*(2), 101–104.
135. Popay, J., Rogers, A., & Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*, *8*(3), 341–351.
136. Timmer, A., Sutherland, L. R., & Hilsden, R. J. (2003). Development and evaluation of a quality score for abstracts. *BMC Medical Research Methodology*, *3*(1), 1–7.
137. Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. *BMJ*, *320*(7226), 50–52.
138. Kendall, M. G., & Smith, B. B. (1939). The problem of m rankings. *Annals of Mathematical Statistics*, *10*(3), 275–287.
139. Lee, L., Packer, T. L., Tang, S. H., & Girdler, S. (2008). Self-management education programs for age-related macular degeneration: A systematic review. *Australasian Journal on Ageing*, *27*(4), 170–176.
140. Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, *28*(3), 763–774.
141. Gearhart, A., Booth, D. T., Sedivec, K., & Schauer, C. (2013). Use of Kendall's coefficient of concordance to assess agreement among observers of very high resolution imagery. *Geocarto International*, *28*(6), 517–526.
142. Pecheva, D., Yushkevich, P., Batalle, D., Hughes, E., Aljabar, P., Wurie, J., Hajnal, J. V., Edwards, A. D., Alexander, D. C., Counsell, S. J., & Zhang, H. (2017). A tract-specific approach to assessing white matter in preterm infants. *Neuroimage*, *157*, 675–694.
143. Filippa, M., Kuhn, P., & Westrup, B. (2017). *Early vocal contact and preterm infant brain development*. Springer International Publishing.
144. Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *Neuroimage*, *221*, 117164.
145. Klungel, O. H., Martens, E. P., Psaty, B. M., Grobbee, D. E., Sullivan, S. D., Stricker, B. H. C., Leufkens, H. G., & de Boer, A. (2004). Methods to assess intended effects of drug treatment in observational studies are reviewed. *Journal of Clinical Epidemiology*, *57*(12), 1223–1231.
146. Rhodes, A. E., Lin, E., & Streiner, D. (1999). Confronting the confounders: The meaning, detection, and handling of confounders in research. *Canadian Journal of Psychiatry*, *44*, 175–179.
147. Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*, *184*, 741–760.
148. Mataix-Cols, D., & Andersson, E. (2021). Ten practical recommendations for improving blinding integrity and reporting in psychotherapy trials. *JAMA Psychiatry*, *78*(9), 943–944.
149. Basoglu, M., Marks, I., Livanou, M., & Swinson, R. (1997). Double-blindness procedures, rater blindness, and ratings of outcome: Observations from a controlled trial. *Archives of General Psychiatry*, *54*(8), 744–748.
150. Agapaki, M., Pinkerton, E. A., & Papatzikis, E. (2022). Music and neuroscience research for mental health, cognition, and development: Ways forward. *Frontiers in Psychology*, *13*, 976883.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Papatzikis, E., Agapaki, M., Selvan, R. N., Pandey, V., & Zeba, F. (2023). Quality standards and recommendations for research in music and neuroplasticity. *Ann NY Acad Sci.*, *1520*, 20–33. https://doi.org/10.1111/nyas.14944