

Effects of Preferred versus Nonpreferred Music on Bench Press Performance:

A Close Replication and Extension Study

Authors: Jasmin C. Hutchinson¹, Jennifer Murphy², Bianca De Lucia¹, Elizabeth O'Neil¹, Diana Curtis¹, Kathleen T. Mellano¹, Luke Pelton¹, and Nicholas Coker¹

¹Department of Exercise Science and Athletic Training, Springfield College, Springfield, MA, USA

²Centre of Applied Science for Health, Technological University Dublin, Dublin, Ireland

Corresponding Author:

Jasmin Hutchinson

263 Alden Street,

Springfield, MA. 01109

(413) 748-3601

jhutchinson@springfieldcollege.edu

Running Head: Bench Press Music Replication

Abstract

Replication is a fundamental aspect of scientific research, yet few replications have been conducted within strength and conditioning. In this paper we attempt to replicate and extend previous research on the effects of preferred (PREF) vs. nonpreferred (NON-PREF) music on bench press performance and motivation using a close replication of a study by Ballmann et al. (2021). The replication sample included 28 resistance-trained men ($M_{\text{age}} = 20.5$ years, $SD = 1.5$), while the extension sample ($n = 47$) comprised resistance-trained men and women ($M_{\text{age}} = 20.6$ years, $SD = 1.5$). Participants listened to PREF and NON-PREF music in a repeated-measures counterbalanced design, while completing bench press repetitions to failure (RTF). Concerning the replication attempt, we found no difference between music conditions for RTF ($p = .545$, Cohen's $d_z = .12$), and the replication and original effect sizes were incompatible ($z = -1.88$, $p = .036$). For motivation there was no difference between music conditions ($p = .084$, $d_z = .34$) and the effect size estimate was incompatible with the original ($z = -4.44$, $p < 0.001$). Thus, the original study findings were not replicated. In the extension study, a two-way ANOVA showed no interaction or main effects of sex and music genre on RTF ($p > .05$). There were, however, main effects of sex ($p = .015$) and genre ($p = .025$) on motivation. In addition, attentional focus was different (more dissociative) for PREF vs. NON-PREF music ($z = -3.11$, $p = .002$), but perceived exertion did not differ between music conditions ($p = 1.00$, $d_z = .00$). Results indicate that music preference does not have a robust effect on bench press performance and associated psychological factors. Athletes, exercisers, and practitioners are encouraged to utilize music that complements the task rather than considering genre preference.

Keywords: Replicability, strength endurance, motivation, music genre

Introduction

The re-examination of published findings in the literature, which is often referred to as ‘reproducibility’ or ‘replication’, is a fundamental aspect of rigorous scientific practice. Although there are no agreed definitions of these terms, ‘reproducibility’ is often referred to as the reanalysis of an existing dataset using the same analysis strategy (Nosek et al., 2022) and ‘replication’ as retesting a claim using the same analyses but with new data (Nosek & Errington, 2020). Although both are important, here we focus our attention on replication. A replication study which supports the original study findings should increase confidence in a claim, otherwise it may decrease confidence in a claim or lead to other lines of research enquiry (Nosek & Errington, 2020). Replication studies are often further categorized as being ‘close’ or ‘extended’. A close replication is a new study that uses the same methods (as close as possible) as the study being replicated. Replication-extension studies not only provide replication evidence but also extend the results of prior studies in new and theoretically important directions (Bonett, 2012).

The replicability of scientific findings has been questioned in recent years, leading to claims of a “crisis of confidence” (Pashler & Wagenmakers, 2012). In psychology, for example, a widespread replication attempt by the Open Science Collaboration (2015) demonstrated that more than half of the empirical findings under scrutiny did not replicate. Increasing interest in replication studies contributes to the active effort to restore credibility to scientific research (Pittelkow et al., 2023), yet replication studies are rare in sport and exercise science (Mesquida et al., 2022).

Several researchers have expressed concern about scientific practices in the field of sport and exercise science and have led a call for an increase in replication studies (Caldwell et al., 2020; Halperin et al., 2018). Factors that affect the replicability of research findings are already

apparent in the field of sport and exercise science, for example, low statistical power, a high positive results rate, and poor data transparency (Borg et al., 2020; Caldwell et al., 2020; Twomey et al., 2021). Given the typical small sample sizes in sport and exercise science (Abt et al., 2020) and the high proportion of positive findings (approximately 81%), there is an unexpected volume of statistically significant findings in the literature for a field that is mostly powered to only observe large effects. Underpowered research designs are concerning as they can increase the proportion of false positive results (Button et al., 2013). Based on the above issues and on tentative evidence for publication bias in the field (Mesquida et al., 2022), it is unsurprising that the replicability of sport and exercise science research has been called into question.

In response to the numerous calls for increased replication in the field, the Sports Science Replication Centre was established (see <https://ssreplicationcentre.com>). The overall aim of this center is to systematically evaluate the replicability of sport and exercise science research. To achieve this aim, replication studies are selected according to a formalized protocol that considers statistical, theoretical and methodological factors (Murphy et al., 2023) and then allocated to laboratories worldwide who are contributing towards this large-scale replication effort. Following this selection protocol, we were allocated a study titled “Effects of Preferred vs. Nonpreferred Music on Resistance Exercise Performance”, published in the *Journal of Strength and Conditioning Research* (Ballmann et al., 2021). This study investigated the effect of listening to preferred (PREF) vs nonpreferred (NON-PREF) genre music on resistance training performance. For this replication, we are specifically interested in the effect of music preference on overall bench press repetitions completed and motivation for the task, and will report this

replication outcome. We did not try to replicate secondary analyses performed in the original study such as power and velocity.

Positive effects of music on athletic performance have been reported previously (see Terry et al., 2020). Carefully selected music can infer an array of psychological, physiological, and psychophysiological benefits that mediate the effect of music on athletic performance. For example, results from a recent meta-analysis indicate significant beneficial effects of music on feeling state, metabolic efficiency, and perceived exertion (Terry et al., 2020). However, most of the existing research has focused on aerobic and endurance-based activities, with evidence of the ergogenic effect of music during resistance exercise being equivocal. Specific to the focus of this study (i.e., muscular endurance), studies have reported increased repetitions to failure (RTF) with self-selected (Bartolomei et al., 2015; Cutrufello et al., 2020) and preferred genre music (Ballmann et al., 2021; Silva et al., 2021). In contrast, Biagini et al. (2012) found no effect of self-selected music on RTF, while Moss et al. (2018) reported a small to moderate effect of music on RTF at low but not high exercise intensities. A clearer understanding of the ergogenic effects of music is necessary for optimizing resistance training regimens for increased performance (Ballmann et al., 2021).

The study assigned to our research group compared the effects of PREF vs NON-PREF music genre on resistance exercise performance. Music preference is an important consideration when examining the effects of music during exercise (Ballmann, 2021). The beneficial effects of music on physiological functioning appear to come from the inherent characteristics of music, regardless of preference (Sleight, 2013), however psychological and psychophysical benefits are usually greater when music preference is taken into account (Terry et al., 2020).

With regard to psychological and psychophysiological effects, music has been shown to increase motivation during resistance training (Ballmann et al., 2020; Ballmann et al., 2021; Lehman et al., 2022), although the sample sizes in these studies were very small ($\text{all} \leq 12$). Rating of perceived exertion (RPE) does not appear to be impacted by music during resistance training (Ballmann et al., 2020; Biagini et al., 2012; Lehman et al., 2022). However, these findings are often confounded by the ergogenic effect of music, that is, an increase in performance without a corresponding increase in perceived exertion is indicative of a positive effect of music on RPE. One of the primary mechanisms by which music provides ergogenic benefits is through attentional disassociation (or distraction) during exercise (Ballmann, 2021). External stimuli, such as music, can draw attention away from internal, fatigue-related cues, thus assuaging effort-related sensations and improving effort tolerance (Hutchinson et al., 2017). Although this effect has been established for aerobic exercise (e.g., Hutchinson & Karageorghis, 2013; Jones et al., 2014), it has yet to be tested in the context of resistance training.

In addition to conducting a close replication of the Ballmann et al. study, we sought to extend this study to address three important considerations. First, almost all prior research on the effects of music during resistance training has focused exclusively on males (Ballmann et al., 2020; Ballmann et al., 2021; Bartolomei et al., 2015; Biagini et al., 2012; Lehman et al., 2022). The current study expanded the participant sample to include females to be more inclusive and to explore potential sex differences. Second, the original study did not assess participants' subjective experience of the music used, presumably because it was assumed that preferred genre would equate to preferred tracks. The current study added a subjective rating of music liking to verify this assumption. Finally, the original study recommended that future investigations should

include measurements of attentional focus and RPE in order to explore these hypothesized mechanisms, therefore these measures were added to the current study.

This study replicated previous work (Ballmann et al., 2021) and tested the alternative hypothesis that listening to PREF music during bench press exercise would result in (a) greater mean performance (RTF) and (b) greater mean motivation, when compared to NON-PREF music. For the additional variables unique to the extension study, it was hypothesized that for attentional focus, the PREF mean would be higher than the NON-PREF mean (i.e., a more dissociative focus), however no difference between conditions was expected for RPE. The interactive effects of sex and music-preference were considered exploratory, and no *a priori* hypothesis was set given the lack of prior research in this area.

Methods

Experimental Approach to the Problem

The present study is a close replication attempt (Brandt et al., 2014), therefore the method was designed to be as similar as possible to the method of the original study. A within-groups study design was used to explore the effects of PREF vs NON-PREF music genre on bench press performance. Experimental conditions were counterbalanced to avoid a potential order effect. This study was pre-registered on the open science framework (<https://osf.io/qm8bf>) and ethics approval was granted by the institution of the first author (approval #1852223). Any deviations from the original study's methodology are transparently reported in the supplementary materials (<https://doi.org/10.17605/OSF.IO/T6ZSU>).

A priori statistical power calculations are typically based on original effect size estimates, yet these estimates are often inflated due to publication bias and low study power (Anderson & Maxwell, 2017). We used the “BUCSS” R package in the statistical software R (Core Team,

2020) which adjusts for publication bias and effect size uncertainty (Anderson et al., 2017). However, this method can be overly conservative; therefore, as per the study selection protocol (Murphy et al., 2023), doubling of the original sample size was used to inform the sample size calculation, resulting in a required sample size of at least 24 participants. An additional four participants were recruited in case of possible loss of data due to attrition or missing values (Peacock & Peacock, 2011). This sample size is greater than that calculated ($n = 19$) by using the original effect size estimate for total repetitions at 95% power with a 5% alpha criterion for a two-tailed test using the “BUCSS” R package in the statistical software R with the BUCSS output at an assurance level of 50%, but less than the assurance level of 80%. Assurance is the percentage of times power would reach or exceed the intended level if the sample-size planning process was to be reproduced many times. Therefore, the calculated sample size of 24 could be considered bias corrected but not uncertainty corrected. See <https://doi.org/10.17605/OSF.IO/T6ZSU> for full calculations and sample size justification.

Participants

Male ($n = 28$) and female participants ($n = 19$) were recruited from a college in the Northeast region of the United States. All participants were informed of the benefits and risks of the investigation prior to signing an institutionally approved informed consent document to participate in the study. The 28 male participants were used as the sample for the replication attempt, while the mixed-sex sample ($n = 47$; $M_{\text{age}} = 20.6$ years, $SD = 1.5$) was used to extend the original study. Inclusion criteria were the same as in the original study. All participants (a) engaged in at least two days per week of resistance training for the past six months, (b) were between the age of 18 and 24 years, (c) were familiar with the bench press exercise, (d) reported

no upper extremity injury in the prior six months, and (e) had no contraindications for exercise as determined by the Physical Activity Readiness Questionnaire (PAR-Q; Warburton et al., 2019).

The male participants used in the replication sample had an average training history of 4.7 ± 1.2 years and reported training an average of 4.7 ± 1.3 days per week. Most participants identified as White (75%), followed by Black or African American (11%), Asian (7%), Hispanic or Latino (3.5%), and mixed race (3.5%). These participants were characteristically similar to the participants in the original study for age and relative bench press performance, although the mean was lower in stature, body mass, and 1RM bench press performance (see Table 1). In the extension sample, participants reported a training history of 4.0 ± 2.4 years and training a mean of 3.9 ± 1.2 days per week. Most participants identified as White (79%), followed by Black or African American (6.5%), mixed race (6.5%), Hispanic or Latino (4%), and Asian (4%).

Procedures

Prior to experimental testing, participants completed the PAR-Q, a demographic information questionnaire, and music preference survey via Qualtrics XM (Provo, Utah, United States). To determine PREF and NON-PREF, participants were asked to rank the following music genres from their most to least favorite: (a) rap, (b) pop, (c) rock, (d) dance/electronic, (e) hip-hop/RnB, and (f) country. For each participant, the selected favorite genre was used for the PREF trial and the selected least favorite genre was used for the NON-PREF trial. Playlists were created using songs from the Billboard Top 10 Singles of 2022 for each genre. Songs were included if they approximated the tempo of the music used in the original study (± 10 bpm). It should be noted that the exact genre categorizations from the original study were unavailable, therefore we selected the closest match to the original study from the lists Billboard had available (see supplementary materials).

Based on the music preference survey, most participants (46.8%) preferred rap music, followed by rock (25.5%), pop (17.0%), dance (6.4%), and hip-hop/RnB (4.3%). No participants selected country as a preferred genre. For non-preferred music genre, most participants selected country (63.8%), followed by dance (17.0%), pop (10.6%), hip-hop/RnB (4.3%), rock (2.1%), and rap (2.1%). Playlists were created in Apple Music and played from an iPad (Apple Inc., Cupertino, California). For each trial, a song was selected randomly from the applicable playlist. The mean tempo for the PREF music selection was 117 ± 11 bpm, and for the NON-PREF music selection was 121 ± 11 bpm.

One-Repetition Maximum Bench Press and Familiarization

Before beginning the 1RM bench press test, participants completed a dynamic warm-up consisting of wall slides, forward/backward arm circles, and arm swings, followed by two sets of four repetitions of bench press using a 20-kg Eleiko Olympic Barbell (Austin, Texas, United States). Following the warm-up, the barbell was loaded to 50% of the participant-reported estimated 1RM and then progressively increased by 2.5-20.0 kg for one repetition until the participant reached failure, which is in line with the original study 1RM protocol. Rest periods between attempts were 3–5 minutes to ensure recovery (Willardson, 2006), and all participants' 1RM was established within four attempts. After 1RM was determined, participants performed a familiarization protocol, for which they were asked to lift the 20-kg Olympic barbell as fast and as explosively as possible for three repetitions, three times. No music was played during the 1RM and familiarization session.

Testing Protocol

The testing protocol consisted of two counterbalanced conditions - PREF and NON-PREF. Trials were completed on different days and separated by a 48-hour washout period.

Trials were scheduled at the same time of day (within participants) to account for any diurnal variation in the dependent variables. Participants were instructed to maintain their typical dietary practices and physical activity levels throughout the intervention, and not to engage in any bench press training for the duration of the study; compliance was verbally confirmed at each visit. During each experimental trial, music was played using Beats Studio 3 (Culver City, California) wireless noise canceling headphones. In the original study, music volume was not controlled, with volume adjusted by the subject to a comfortable level (Ballmann et al., 2021, p. 1652). Given that music volume is a variable that can affect arousal and exercise performance (Hutchinson & Jones, 2020), we chose to standardize the music volume at 75dB for all participants. The bench press exercise was completed using a standard 20-kg Ivanki powerlifting barbell (San Pedro, California), Samson half rack (Las Cruces, New Mexico), and Powertec Streamline bench (Paramount, California).

For each trial, participants began with a brief warm-up that consisted of 30% 1RM for five repetitions followed by one set of 50% of 1RM for three repetitions. After five minutes of rest, the barbell was loaded with 75% of 1RM. After the participant placed the headphones on and the music started, participants immediately lifted 75% of their previously obtained 1RM for as many repetitions as possible, as explosively as possible, until failure. Failure was defined as a participant verbally indicating they cannot continue or when the participant was unable to complete a repetition due to momentary muscular failure (Fisher et al., 2011) resulting in downward movement of the barbell during the concentric portion of the lift. The participant's head, hips, and upper back remained in contact with the bench throughout the movement with feet flat on the floor. Each repetition was executed with a full eccentric (lower the bar to lightly touch the mid chest) and concentric phase (press to full elbow extension directly over the

shoulders) and without bouncing the barbell off the chest. Any lift that deviated from this correct bench press form was not counted toward the total (Ballmann et al., 2021); however, no deviations were noted.

Once each trial was complete, participants immediately reported their motivation, attentional focus, and RPE. As in the original study, motivation was assessed using a visual analogue scale with a 100-mm line, anchored at either end with the descriptors of ‘least motivated’ to ‘most motivated’. Participants marked on the scale how motivated they felt during the lift with a vertical line. Motivation scores were obtained by measuring the distance, in millimeters, from zero to the vertical line marked by the participant using a ruler. Attentional focus was assessed using a single-item attention scale (Tammen, 1996). Scale instructions were as follows: “To what extent was your attentional focus during the task internal (i.e., related to your body) or external (i.e., related to the outside environment)?” The scale ranges from “0” which represents an internal focus (e.g., heart rate, muscle fatigue, and breathing) to “100” representing an external focus (e.g., daydreaming, environment, and music). Participants circled a number on the attention scale that best represented their focus during the lift. Higher scores indicate greater attentional dissociation. RPE was assessed by means of Borg’s RPE scale (Borg, 1998). Scaling ranged from 6 (“no exertion”) to 20 (“maximal exertion”) and participants were asked to indicate the number that corresponded to their perceived exertion during the lift.

Statistical Analyses

Data were analyzed using R (version 4.2.1; R Core Team, 2020). Paired-sample t-tests were used to compare bench press performance (i.e., RTF) and motivation for the male only replication sample. Paired-sample t-tests were also used to statistically compare RPE, attentional focus, and song rating data between conditions for the mixed-sex sample. Where data failed

normality checks ($p < 0.05$), a Wilcoxon signed rank test was conducted. Separate two-way mixed analyses of variance (ANOVA) were used to statistically analyze the interaction between sex and music preference on bench press performance and motivation. If Mauchly's test indicated violations of the sphericity assumption, Greenhouse-Geisser adjustments were made to the relevant F test. Following any significant effects, Bonferroni post hoc pairwise comparisons were used.

To evaluate the replication outcome, the following criteria were used: for the original study to be considered replicated, the replication effect must also be statistically significant and in the same direction as the original effect (Murphy et al., 2023). We also compared the original and replication effect size estimates using the *TOSTER* R package, (version 0.5.0; Caldwell, 2022). Simply, each effect size estimate was transformed into a z-score and a z-test was conducted. The significance level for all analyses was set at $\alpha < 0.05$. The raw data and code for the analyses can be found at <https://doi.org/10.17605/OSF.IO/T6ZSU>.

Results

Prior to conducting analyses, the data were inspected for missing values, outliers, and basic assumptions. No missing values were present in the data. For the replication study sample, the dependent variables and paired differences were all normally distributed as assessed by the Shapiro Wilk test ($p > .01$). In the mixed-sex extension study sample, five outlying scores were noted in the total repetitions data, one of which was extreme. The extreme score was corrected by adjusting to the next highest or lowest value not considered an outlier (Tabachnick et al., 2019). The other outliers were not extreme outliers, and the data was normal, therefore, these scores remained unadjusted. All dependent variables were normally distributed for the extension study sample, with the exception of the attentional focus data ($W = .934, p < .001$). Table 2

compares means and standard deviations between the original and replication studies. Table 3 shows means and standard deviations for all dependent variables for the mixed-sex sample.

Replication Results

For the replication attempt, paired sample t-tests were used to determine if there was a statistically significant mean difference in the total number of bench press RTF and motivation between PREF vs. NON-PREF music conditions. There was no significant difference between conditions on the total number of RTF, $t(27) = 0.61, p = .545, M_{\text{diff}} = 0.18, 95\% \text{ CI } [-0.42, 0.78]$, Cohen's $d_z = 0.12 [-0.26, 0.49]$. Likewise, there was no significant difference in motivation between music conditions, $t(27) = 1.79, p = .084, M_{\text{diff}} = 9.73, 95\% \text{ CI } [-1.41, 20.86]$, Cohen's $d_z = 0.34, 95\% \text{ CI } [-0.05, 0.72]$.

The original study observed a significant difference between music conditions for total RTF ($p = .005$) but the replication study did not ($p = .545$). The replication study effect size estimate for the total repetitions was Cohen's $d_z = 0.12$ compared to the original which was Cohen's $d = 0.84$. According to the z-test, the replication and original effect size estimates were incompatible [$z = -1.88, p = .036, \text{ difference} = 0.72$]. Concerning the evaluation of our replication attempt, we consider the original effect size reported by Ballmann et al. (2021) not replicated, albeit this should be interpreted with caution due to statistical noise (see discussion).

The original study observed a significant difference between music conditions for motivation ($p < 0.001$) but the replication study did not ($p = .084$). The replication study effect size estimate was Cohen's $d_z = 0.34$ compared to the original which was Cohen's $d = 5.90$. According to the z-test, the replication and original effect size estimates were not compatible [$z = -4.44, p < 0.001, \text{ difference} = 5.56$], therefore, the original effect size reported by Ballmann et al., (2021) was not replicated.

Extension Results

Two-way mixed ANOVAs were conducted to explore the interactive effects of sex and music preference on bench press performance (i.e., RTF) and motivation in a larger, mixed-sex sample ($n = 47$). The interaction of sex and music preference on the total number of RTF was not significant, $F(1, 45) = 0.00$, $p = .982$, $\eta_p^2 < .001$, nor any main effect of sex $F(1, 45) = .42$, $p = .521$, $\eta_p^2 = .009$, nor main effect of music preference $F(1, 45) = .14$, $p = .713$, $\eta_p^2 = .003$. There was also no significant interaction of sex and music preference on motivation, $F(1, 45) = .16$, $p = .692$, $\eta_p^2 = .004$, but there was a significant main effect of sex $F(1, 45) = 6.47$, $p = .015$, $\eta_p^2 = .126$, and a significant effect of music preference $F(1, 45) = 5.35$, $p = .025$, $\eta_p^2 = .106$. Bonferroni-adjusted comparisons indicated that males reported greater motivation for the task than females ($M_{\text{diff}} = 12.25$, 95% CI [2.55, 21.96], $p = .015$, Cohen's $d_s = 0.76$, 95% CI [0.15, 1.35]). PREF music also resulted in greater motivation than NON-PREF music ($M_{\text{diff}} = 11.76$, 95% CI [1.52, 21.99], $p = .025$, Cohen's $d_z = 0.34$, 95% CI [0.04, 0.63]).

Three additional tests were performed on attentional focus, RPE, and song rating which were unique to this study. The first compared the effect of listening to PREF vs NON-PREF music on attentional focus. The data for attentional focus was non-normal ($W = .934$, $p = .011$), therefore a non-parametric statistical analysis was used. The Wilcoxon signed rank test was significantly different for PREF compared to NON-PREF, $z = -3.11$, $p = .002$, with higher (i.e., more dissociative) attentional focus in the PREF music condition. A paired t-test compared the effect of listening to PREF vs NON-PREF music on RPE; there was no significant difference between conditions on RPE, $t(46) = 0.00$, $p = 1.00$, $M_{\text{diff}} = 0.00$, 95% CI [-0.61, 0.61], Cohen's $d_z = 0.00$, 95% CI [-0.29, 0.29]. Finally, the song ratings for PREF and NON-PREF were also compared as a manipulation check using a paired t-test. As expected, there was significantly

higher rating for PREF compared to NON-PREF, $t(46) = 3.52, p < .001, M_{\text{diff}} = 2.10, 95\% \text{ CI } [0.90, 3.30]$, Cohen's $d_z = 0.51 [0.21, 0.82]$.

Discussion

The primary aim of this study was to determine if we could replicate the observed effects of PREF vs NON-PREF music on bench press performance and motivation using a close replication of an original study (Ballmann et al., 2021). In addition, the present study sought to extend the original research by using a mixed-sex sample and examining the effect of listening to PREF vs NON-PREF music on attentional focus and RPE.

Contrary to the findings of the original study, the number of bench press repetitions were similar across music conditions in the current study for both the male-only sample and the mixed-sex sample. Moreover, we failed to reject the null hypothesis for motivation in the replication sample, although there was a difference in the mixed-sex sample. Using the parameters of a 'successful replication' (Brandt et al., 2014; Murphy et al., 2023), we were unable to replicate Ballmann et al.'s original results from a null hypothesis significance testing perspective for the replication sample. Our z-test showed that the replication and original effect size estimates were incompatible ($p = .036$), with a large difference in the standardized values (0.72). Furthermore, when considered with other criteria, the original and replication effect size estimates were not in the same direction; the mean for total repetitions was lower in the PREF group compared to the NON-PREF group which is in contrast with the original study. The original effect size estimate did fall into the confidence intervals of the replication effect size, however, this is not entirely informative due to the imprecision of the confidence intervals (95% CI: -0.26, 0.49); this is an identified issue with this method as discussed by Asendorpf et al., (2013). Based on all of the above outcomes, we consider the original findings not replicated for

total repetitions. The small sample size of the original study suggests the investigation was likely underpowered, which reduces the likelihood that a statistically significant result reflects a true effect (Button et al., 2013). Similarly, this replication study is likely underpowered to detect the plausible population effect size, or even provide a precise estimate of this effect. This should therefore be considered with all outcomes presented below.

Given that most prior studies used all-male samples, the lack of difference in bench press performance regardless of sex, is important to note. However, it should be considered that this was an exploratory analysis and our study was not powered to detect sex differences. The only other located study that used a mixed-sex sample reported no sex differences in bench press performance when comparing music and no-music conditions (Cutrufello et al., 2020). Ours is the first study to compare PREF to NON-PREF music, but findings are consistent and suggest no difference in the influence of music preference between men and women on performance outcomes in a resistance training context. This contrasts with findings from other exercise modalities; for example, Cole and Maeda (2015) reported that PREF (vs. NON-PREF) music had a greater effect on the endurance running performance of women than men. In this case, the authors attributed the difference to the observation that “women are known to pay closer attention to the rhythmical qualities of music than men” (p.395); a mechanism that was absent in the current study given the lack of sensorimotor synchronization during the activity.

When assessing the inconsistent effects of preferred music on performance, the intensity and duration of the task itself warrants consideration. A bench press to fatigue at 75% 1RM typically takes less than 30 seconds to complete, which may provide insufficient exposure to the music for preference to have an effect. Most popular music tracks have an introduction ~15 seconds, therefore the more motivational parts of the song, such as the chorus and pre-chorus

(Beall, 2009), will not be heard during the bench press task. Longer duration tasks have prolonged exposure to music and are also not as reliant on the impact of a single track. The current study used a load of 75% of the 1RM, as did others (Ballmann et al., 2020; Biagini et al., 2021; Lehman et al., 2022). However, Bartolomei et al. (2015) used a lower percentage of 1RM (60%) while Moss et al. (2018) compared a range of intensities from 30% to 80% 1RM. Differences in load could influence the repetition cadence, thereby altering the rhythm of the movement. Music is more likely to elicit auditory-motor unit synchronization if the rhythm of the music and task are closely aligned. This has the potential to aid in metabolic efficiency via improved neuromuscular and kinetic efficiency (Terry et al., 2020). Future research should try to match the music tempo with the motor task to better elucidate the influence of music on performance.

Another methodological consideration is the process of song selection. Self-selected music may be more influential than preferred music genre. Genre categorization is imprecise and of declining importance in modern music (Silver et al., 2016). Moreover, there are an increasing number of crossover artists who blend multiple music genres (Shi et al., 2018) and whose work does not fall neatly into one genre category or another. During our music selection process, it was noted that many songs appeared on the Billboard Top 10 Singles of 2022 for more than one genre. In fact, six of the top 10 rap tracks were also listed in the top 10 hip-hop R & B tracks. Although overall song ratings in the present study were significantly different between conditions, there was notable interindividual variability in the ratings. In the replication sample, for example, eight of 28 participants (29%) actually rated the NON-PREF song higher than the PREF song. We contend that a more careful music selection protocol is needed to adequately test the research question. Moreover, future research ought to consider examining more sophisticated

musical elements, such as complexity, accents, and beat perception that may be more important than genre for performance outcomes.

In the current study, listening to preferred music resulted in higher motivation scores only in our larger mixed-sex sample. Ballman et al.'s findings for the all-male sample were not replicated. Notably, the effect size estimate was considerably lower in our extension sample ($d_z = 0.34$) than in the original study ($d = 5.90$). Although these effect sizes were calculated differently, lower sample sizes reduce statistical power, leading to an overestimation of the effect size estimate (Button et al., 2013). The small sample size in the original study ($n = 12$) likely resulted in an effect size estimate that did not reflect the true effect at a population level. The larger sample size in our extension ($n = 47$) study may provide a more accurate estimation of the effect of music preference on motivation for bench press exercise than the original effect size; however, we believe our effect size is still imprecise due to the width of the confidence interval (95% CI: -0.05, 0.72). Although preferred music did result in higher motivation scores in the mixed-sex sample, bench press performance was not positively influenced. One reason for the disconnect between motivation and performance is the short duration of the task which, as previously mentioned, likely provided insufficient exposure to the music for any difference in motivation to have a discernible effect on performance.

In addition to attempting to replicate the outcome measures of Ballman et al. (2021), the present study also assessed RPE and attentional focus in order to explore these potential mechanisms of effect. No difference in RPE was found between conditions, which is consistent with prior research (Ballmann et al., 2020; Biagini et al., 2012; Lehman et al., 2022). Attentional focus was more dissociative when listening to PREF music, compared to NON-PREF. To the best of our knowledge, this is the first study to examine the effect of music preference on

attentional focus during a physical task. That PREF captures attention to a greater extent than NON-PREF music is perhaps not surprising. Preferred music is likely to be more familiar to the listener, and music familiarity plays an important role in the emotional engagement and brain activation of listeners (Pereira et al., 2011). Moreover, there may be an active attempt on the part of the listener to ‘tune out’ music that is disliked. Typically, a more dissociative attentional focus is associated with lower RPE during exercise (Terry et al., 2020), but this association appears more prevalent in prolonged endurance tasks (e.g., distance running or cycling) where exertion tolerance plays a larger role in determining performance. Again, the short duration of the task in the present study likely accounts for the observed disconnect between attentional focus and RPE.

In conclusion, by conducting a close replication of Ballmann et al. (2021) we aimed to assess the robustness of the reported significant effect of PREF music over NON-PREF music on bench press performance and motivation. We were unable to replicate the findings from the original study when using experimental procedures that matched the original as closely as possible, albeit with larger sample sizes with the aim to increase statistical power. The estimates were incompatible in the z-test and the replication effect size estimate was considerably smaller than the original. However, any effect size estimates from this replication should be considered with caution due to the large width of the confidence intervals.

Table 1*Mean (SD) Descriptive Characteristics for Participants in the Original and Replication Studies*

	Original Study Males (<i>n</i> = 12)	Current Study Males (<i>n</i> = 28)	Current Study Females (<i>n</i> = 19)	Current Study Total Sample (<i>N</i> = 47)
Age (years)	20.5 (1.2)	20.5 (1.5)	20.4 (1.1)	20.6 (1.5)
Height (cm)	182.8 (6.8)	174.8 (6.1)	164.2 (6.6)	170.5 (8.2)
Body mass (kg)	95.1 (17.8)	83.3 (11.9)	70.0 (16.1)	77.9 (15.0)
Bench press, 1-RM (kg)	127.9 (40.2)	107.3 (22.5)	48.5 (10.0)	83.5 (34.2)
Relative bench press, (1RM [kg]/BM [kg])	1.35 (0.1)	1.29 (0.2)	0.70 (0.2)	1.05 (0.3)

Note: 1RM = 1 repetition maximum.

Table 2*Mean (SD) of Dependent Variables between the Original and Replication Studies*

	Original Study		Replication Study	
	Preferred Music	Nonpreferred Music	Preferred Music	Nonpreferred Music
Total repetitions (AU)	10.58 (2.07)	8.90 (1.80)	13.07 (2.12)	13.25 (2.14)
Motivation (AU)	80.40 (11.20)	18.8 (9.29)	71.93 (19.02)	62.20 (25.07)

Note: AU = Arbitrary unit.

Accepted: In Press

Table 3*Mean (SD) Descriptive Statistics for Mixed-Sex Sample Tests*

	Preferred Music	Nonpreferred Music
Total repetitions	12.87 (3.10)	13.04 (2.91)
Motivation (AU)	67.80 (21.01)	56.44 (26.84)
Attentional focus (AU)	49.04 (27.50)	35.11 (22.75)
RPE (AU)	15.31 (2.62)	15.31 (2.99)
Song rating (AU)	6.52 (2.88)	4.41 (3.05)

Note: RPE = Rating of perceived exertion. AU = Arbitrary unit.

Accepted: In Press

References

- Abt, G., Boreham, C., Davison, G., Jackson, R., Nevill, A., Wallace, E., & Williams, M. (2020). Power, precision, and sample size estimation in sport and exercise science research. *Journal of Sports Sciences*, 38(17), 1933–1935.
<https://doi.org/10.1080/02640414.2020.1776002>
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28(11), 1547–1562.
<https://doi.org/10.1177/0956797617723724>
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the “replication crisis”: Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52(3), 305–324. <https://doi.org/10.1080/00273171.2017.1289361>
- Asendorpf, J.B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J.J., Fiedler, K., Fiedler, S., Funder, D.C., Kliegl, R., Nosek, B.A. & Perugini, M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27(2), 108–119.
<https://doi.org/10.1002/per.1>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454.
<https://doi.org/10.1038/533452a>
- Ballmann, C. G. (2021). The influence of music preference on exercise responses and performance: A review. *Journal of Functional Morphology and Kinesiology*, 6(2), 33.
<https://doi.org/10.3390/jfmk6020033>
- Ballmann, C. G., Cook, G. D., Hester, Z. T., Kopec, T. J., Williams, T. D., & Rogers, R. R. (2020). Effects of preferred and non-preferred warm-up music on resistance exercise performance. *Journal of Functional Morphology and Kinesiology*, 6(1), 3.
<https://doi.org/10.3390/jfmk6010003>
- Ballmann, C. G., McCullum, M. J., Rogers, R. R., Marshall, M. M., & Williams, T. D. (2021). Effects of preferred vs. nonpreferred music on resistance exercise performance. *Journal of Strength and Conditioning Research*, 35(6), 1650–1655.
<https://doi.org/10.1519/JSC.0000000000002981>

- Bartolomei, S., Michele, R. D., & Merni, F. (2015). Effects of self-selected music on maximal bench press strength and strength endurance. *Perceptual and Motor Skills*, 120(3), 714–721. <https://doi.org/10.2466/06.30.PMS.120v19x9>
- Beall, E. (2009). *The Billboard guide to writing and producing songs that sell: How to create hits in today's music industry* (1st ed.). Billboard Books, Watson-Guptill Pub.
- Biagini, M. S., Brown, L. E., Coburn, J. W., Judelson, D. A., Statler, T. A., Bottaro, M., Tran, T. T., & Longo, N. A. (2012). Effects of self-selected music on strength, explosiveness, and mood. *Journal of Strength and Conditioning Research*, 26(7), 1934–1938. <https://doi.org/10.1519/JSC.0b013e318237e7b3>
- Bonett, D. G. (2012). Replication-extension studies. *Current Directions in Psychological Science*, 21(6), 409–412. <https://doi.org/10.1177/0963721412459512>
- Borg, D. N., Bon, J. J., Sainani, K. L., Baguley, B. J., Tierney, N. J., & Drovandi, C. (2020). Comment on: 'Moving sport and exercise science forward: A call for the adoption of more transparent research practices.' *Sports Medicine*, 50(8), 1551–1553. <https://doi.org/10.1007/s40279-020-01298-5>
- Borg, G. (1998). *Borg's perceived exertion and pain scales*. Human Kinetics.
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & Van 't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Caldwell, A. R. (2022). Exploring equivalence testing with the updated TOSTER R package. *PsyArXiv*. <https://doi.org/10.31234/osf.io/ty8de>
- Caldwell, A. R., Vigotsky, A. D., Tenan, M. S., Radcl, R., Mellor, D. T., Kreutzer, A., Lahart, I. M., Mills, J. P., & Boisgontier, M. P. (2020). Moving sport and exercise science forward: A call for the adoption of more transparent research practices. *Sports Medicine*, 50(3), 449–459. <https://doi.org/10.1007/s40279-019-01227-1>

- Cole, Z., & Maeda, H. (2015). Effects of listening to preferential music on sex differences in endurance running performance. *Perceptual and Motor Skills*, 121(2), 390–398. <https://doi.org/10.2466/06.PMS.121c20x9>
- Cutrufello, P. T., Benson, B. A., & Landram, M. J. (2020). The effect of music on anaerobic exercise performance and muscular endurance. *The Journal of Sports Medicine and Physical Fitness*, 60(3). <https://doi.org/10.23736/S0022-4707.19.10228-9>
- Fisher, J., Steele, J., Bruce-Low, S., & Smith, D. (2011). Evidence-based resistance training recommendations. *Medicina Sportiva*, 15(3), 147–162. <https://doi.org/10.2478/v10036-011-0025-x>
- Halperin, I., Vigotsky, A. D., Foster, C., & Pyne, D. B. (2018). Strengthening the practice of exercise and sport-science research. *International Journal of Sports Physiology and Performance*, 13(2), 127–134. <https://doi.org/10.1123/ijsp.2017-0322>
- Hutchinson, J. C., & Jones, L. (2020). Affect and music. In *The routledge international encyclopedia of sport and exercise psychology* (Vol. 2, pp. 21–36). Routledge.
- Hutchinson, J. C., & Karageorghis, C. I. (2013). Moderating influence of dominant attentional style and exercise intensity on responses to asynchronous music. *Journal of Sport and Exercise Psychology*, 35(6), 625–643. <https://doi.org/10.1123/jsep.35.6.625>
- Hutchinson, J. C., Karageorghis, C. I., & Black, J. D. (2017). The diabeates project: Perceptual, affective and psychophysiological effects of music and music-video in a clinical exercise setting. *Canadian Journal of Diabetes*, 41(1), 90–96. <https://doi.org/10.1016/j.cjcd.2016.07.009>
- Jones, L., Karageorghis, C. I., & Ekkekakis, P. (2014). Can high-intensity exercise be more pleasant? Attentional dissociation using music and video. *Journal of Sport and Exercise Psychology*, 36(5), 528–541. <https://doi.org/10.1123/jsep.2013-0251>
- Lehman, J. T., Whitmire, B. G., Rogers, R. R., Williams, T. D., & Ballmann, C. G. (2022). Effects of respite music on repeated upper-body resistance exercise performance. *International Journal of Exercise Science*, 15(7), 79–87.
- Mesquida, C., Murphy, J., Lakens, D., & Warne, J. (2022). Replication concerns in sports and exercise science: A narrative review of selected methodological issues in the field. *Royal Society Open Science*, 9(12), 220946. <https://doi.org/10.1098/rsos.220946>

- Moss, S. L., Enright, K., & Cushman, S. (2018). The influence of music genre on explosive power, repetitions to failure and mood responses during resistance exercise. *Psychology of Sport and Exercise*, 37, 128–138. <https://doi.org/10.1016/j.psychsport.2018.05.002>
- Murphy, J., Mesquida, C., Caldwell, A. R., Earp, B. D., & Warne, J. P. (2023). Proposal of a selection protocol for replication of studies in sports and exercise science. *Sports Medicine*, 53(1), 281–291. <https://doi.org/10.1007/s40279-022-01749-1>
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, 18(3), e3000691. <https://doi.org/10.1371/journal.pbio.3000691>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530. <https://doi.org/10.1177/1745691612465253>
- Peacock, J.L., & Peacock, P.J. (2011). *Oxford handbook of medical statistics*. Oxford University Press.
- Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: Familiarity matters. *PLoS ONE*, 6(11), e27241. <https://doi.org/10.1371/journal.pone.0027241>
- Pittelkow, M. M., Field, S. M., Isager, P. M., van't Veer, A. E., Anderson, T., Cole, S. N., Dominik, T., Giner-Sorolla, R., Gok, S., Heyman, T., Jekel, M., Luke, T. J., Mitchell, D. B., Peels, R., Pendrous, R., Sarrazin, S., Schauer, J. M., Specker, E., Tran, U. S., ... Van Ravenzwaaij, D. (2023). The process of replication target selection in psychology: What to consider? *Royal Society Open Science*, 10(2), 210586. <https://doi.org/10.1098/rsos.210586>
- R Core Team (2020). *R: A language and environment for statistical computing* [Computer software].

- Shi, Y., Lim, Y., & Suh, C. S. (2018). Innovation or deviation? The relationship between boundary crossing and audience evaluation in the music field. *PLOS ONE*, 13(10), e0203065. <https://doi.org/10.1371/journal.pone.0203065>
- Silva, N. R. D. S., Rizardi, F. G., Fujita, R. A., Villalba, M. M., & Gomes, M. M. (2021). Preferred music genre benefits during strength tests: Increased maximal strength and strength-endurance and reduced perceived exertion. *Perceptual and Motor Skills*, 128(1), 324–337. <https://doi.org/10.1177/0031512520945084>
- Silver, D., Lee, M., & Childress, C. C. (2016). Genre complexes in popular music. *PLOS ONE*, 11(5), e0155471. <https://doi.org/10.1371/journal.pone.0155471>
- Sleight, P. (2013). Cardiovascular effects of music by entraining cardiovascular autonomic rhythms music therapy update: Tailored to each person, or does one size fit all? *Netherlands Heart Journal*, 21(2), 99–100. <https://doi.org/10.1007/s12471-012-0359-6>
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Tammen, V. V. (1996). Elite middle and long distance runners associative/dissociative coping. *Journal of Applied Sport Psychology*, 8(1), 1–8. <https://doi.org/10.1080/10413209608406304>
- Terry, P. C., Karageorghis, C. I., Curran, M. L., Martin, O. V., & Parsons-Smith, R. L. (2020). Effects of music in exercise and sport: A meta-analytic review. *Psychological Bulletin*, 146(2), 91–117. <https://doi.org/10.1037/bul0000216>
- Twomey, R., Yingling, V., Warne, J., Schneider, C., McCrum, C., Atkins, W., Murphy, J., Romero Medina, C., Harlley, S., & Caldwell, A. (2021). Nature of our literature: A registered report on the positive result rate and reporting practices in kinesiology. *Communications in Kinesiology*, 1(3). <https://doi.org/10.51224/cik.v1i3.43>
- Warburton, D. E. R., Jamnik, V., Bredin, S. S. D., Shephard, R. J., & Gledhill, N. (2019). The 2020 Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) and electronic Physical Activity Readiness Medical Examination (ePARmed-X+). *The Health & Fitness Journal of Canada*, 58–61. <https://doi.org/10.14288/HFJC.V12I4.295>
- Willardson, J., C. (2006). A brief review: Factors affecting the length of the rest interval between resistance exercise sets. *Journal of Strength and Conditioning Research*, 20(4), 978–984. <https://doi.org/10.1519/R-17995.1>

Additional Information

Data Accessibility

The supplementary materials, data and code for the analysis can be found at <https://doi.org/10.17605/OSF.IO/T6ZSU>

Author Contributions

Contributed to conception and design: JCH, JM, NC

Contributed to acquisition of data: BD, EO, DC, KTM, LP, NC

Contributed to analysis and interpretation of data: JCH, JM, LP

Drafted and/or revised the article: JH, JM, BD, EO, DC, KTM, LP

Approved the submitted version for publication: JCH, JM

Funding

None.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. We wish to disclose that Jenny Murphy is the current Outreach Chair for the Society of Transparency, Openness, and Replication in Kinesiology (STORK), but was not involved in any aspect of the editorial handling of this manuscript, except as a co-author.

Acknowledgements

The authors would like to thank Kaitlyn Clouse, Garrett Hillard, Clayton Knibbs, and Dominic Velazquez for their assistance with data collection, and the participants who volunteered their time for this study.