

An Exploratory Analysis of Self, Other, and Computational Measures of Creativity in Song Lyric Writing

Music & Science
Volume 8: 1–16© The Author(s) 2025
DOI: 10.1177/20592043251364625
journals.sagepub.com/home/mns

Andrea Schiavio^{1,*} , Adrian Kempf^{2,*} , Freya Bailes³
and Mathias Benedek²

Abstract

The present study offers an exploratory analysis of the creative processes and products at the heart of song lyric writing. Drawing on insights from creative metacognition research and work that integrates individual and collective aspects of creativity, we focus on instances where individuals generate artistic outcomes both for themselves and for others. Sixty-three participants (35 musical experts and 28 novices) were invited to complete a series of questionnaires and craft new lyrics for two distinct musical excerpts. They were instructed to write for either themselves (envisioning they would sing the final product) or for another individual (considering someone else would perform). The creativity of the lyrics was assessed using three complementary approaches: (i) self-ratings by participants, (ii) qualitative evaluations by six expert raters, and (iii) algorithmic analysis through distributional semantic modeling. Although we did not find significant differences in creativity between lyrics created for oneself and lyrics created for others, our analysis revealed that participants' behaviors and states are more predictive of self-assessment and computational ratings of creativity than expert ratings. Furthermore, we identified a robust correlation between expert ratings and computational evaluations of the lyrics' creativity.

Keywords

Creativity, creativity assessment, metacognition, musical creativity, music cognition

Submission date: 18 December 2024; Acceptance date: 24 June 2025

Introduction

Creativity is a fundamental aspect of human cognition and one that has long been a subject of inquiry across a range of scientific and artistic domains. Much research on creativity gravitates around two interconnected issues, namely the evaluation of what makes a product “creative” and the understanding of the cognitive and neural processes involved in creative effort (Abraham, 2018; Benedek et al., 2019). As definitions and understandings of creativity exhibit different properties across cultures and historical periods (Niu & Sternberg, 2006), both orientations necessarily involve a focus on the relationships between single individuals and the society in which they are embedded (Glăveanu, 2013). Additionally, they entail the study of *creative metacognition* – that is, thinking about one's (creative) thinking that involves the monitoring and regulation of the

ongoing creative process (see Jia et al., 2019). One way to investigate the interplay among these key factors of creative thought and action is to examine how the metacognitive dimension of self-perceived creativity influences the

¹ School of Arts and Creative Technologies, University of York, York, UK

² Institute of Psychology, University of Graz, Graz, Austria

³ School of Music, University of Leeds, Leeds, UK

*Contributed equally

Corresponding author:

Mathias Benedek, Institute of Psychology, University of Graz,
Universitätsplatz 2, 8010 Graz, Austria.

Email: mathias.benedek@uni-graz.at

Data Availability Statement included at the end of the article



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page

(<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

relationship between creative evaluation and production, and how this relationship may be affected when the social context (i.e., the spectrum of individuality and collectivity) is manipulated. In other words, research should investigate how social expectations and external judgments shape metacognitive evaluation processes, with particular attention to the alignment or discrepancy between external assessments – such as those made by experts or through computational methods (e.g., automated creativity scoring) – and individuals' self-evaluations. This comparison between evaluative perspectives can also illuminate how these forms of assessment may rely on different levels of expertise.

To address these *desiderata*, the present work reports on an original empirical study carried out in a musical context – as we shall see, an ideal context to explore creativity from a range of complementary perspectives (e.g., Antović et al., 2023; Kempf et al., 2024a; 2024b; Schiavio et al., 2024). Arguably, no musical activity can be considered separate from the creative processes that underpin it – whether it involves learning to play an instrument, improvising, composing, or even listening to music (Cook, 2018, Schiavio et al., 2022c). Lyric generation is a particularly interesting case of artistic creativity. Throughout history, individuals with diverse musical backgrounds¹ have demonstrated an ability to blend instrumental music with lyrical content that is *original* and contextually *appropriate* – two fundamental aspects of creativity (see Runco & Jaeger, 2012). From an empirical standpoint, this setting offers significant advantages, as participants may feel less intimidated by the process of creating lyrics compared to performing music, which can often evoke heightened levels of anxiety or self-consciousness in both novice and expert performers. However, while the distinction between the products and processes of creativity in music has been increasingly addressed in recent literature (see van der Schyff et al., 2018), framing individuality, collectivity, and metacognition in relation to creativity and musical creativity may require deeper analysis – especially when these concepts are tied to the notion of “assessment.” Therefore, in the following section we will explore individual and collective perspectives on creativity and metacognition in greater depth before introducing the main study. As we will see, framing an empirical task where participants consider either themselves or someone else as the central actor in a creative behavior may offer valuable insights into how these two aspects compare, offering an opportunity to explore the potential for forms of creativity rooted in intersubjectivity to manifest even within a single individual.

Understanding Creativity

Understanding human creativity requires navigating its complex experiential landscape, including examining the overlaps and distinctions between individual and collective aspects of creative thought and action. This distinction has been the focus of previous research on musical improvisation and creativity (Schiavio et al., 2022a; 2022b; 2024)

as well as studies in skill acquisition, which increasingly suggest that even seemingly individual activities often involve a “hidden” intersubjective dimension (Schiavio et al., 2019; see also Høffding & Satne, 2021). The core idea is that musical skills are developed and refined within a “community of practice” (Lave & Wenger, 1991), making social and cultural engagement fundamental to their emergence and expression, even in individual contexts. Accordingly, musicians, performers, lyricists, or composers, should not be viewed as isolated creators. This challenges the traditional myth of the lone genius, which has historically placed undue emphasis on individual agency in creativity research (see Cook, 2018; Hill, 2018). Instead, creative ideas and choices arise through continuous interaction with cultural and social environments. However, capturing this dynamic is complex, as it unfolds at multiple levels of awareness: While some creators consciously recognize the influences shaping their work, others may be less aware of or able to articulate these external factors.

It should also be said that individual and collective understandings of creative phenomena have provided important advances to our comprehension of creativity via empirical and theoretical contributions (e.g., Benedek et al., 2020; Feist, 1998; Simonton, 1999). However, seminal accounts of individual creativity have focused on the role of divergent and convergent thinking (Guilford, 1950; Mednick, 1968; Razumnikova, 2013), as well as attentional capacity (Mendelsohn, 1976; Kasof, 1997), aiming to offer systematic characterizations of how creative ideas emerge via processes of problem solving and (re-) combination of existing concepts. Conversely, scholarship that explores the networked realization of creative products and ideas has highlighted the communicative, ecological, organizational, and participatory resources of collective thought and action (Amabile, 1996; Perry-Smith & Challey, 2003; Sawyer & De Zutter, 2009). For both orientations, a wide range of factors is carefully considered to articulate generalizations that inform how different cognitive mechanisms operate and support creative ideation.

Divergent thinking tasks, for example, are often associated with combinations of ideas, whose properties (e.g., coherence, originality, fluency, and flexibility) can be tested with well-known assessment batteries and scoring systems (Runco & Acar, 2012; Saretzki et al., 2024). Research on individual creativity has also adopted a neuroscientific approach; this can offer important insights into the main brain networks associated with various creative abilities, including divergent thinking (Runco & Yoruk, 2014). Studies on collective, or distributed, forms of creativity, instead, often assume that crucial aspects of creative thought and action may be missed in individual accounts (Mockros & Csikszentmihalyi, 1999; Sawyer, 2006, 2012) and therefore tend to emphasize the importance of collaborative processes in understanding the full scope of creative dynamics. Empirical investigations driven by a similar rationale often employ various methodologies, including the use of scales (Silvia et al., 2012), focus group interviews

(e.g., Bissola & Imperatori, 2011), and quantitative analyses of behavioral parameters at the group level (e.g., the observed changes in movement coordination between pairs of pianists improvising together reported in Walton et al., 2017).

This constellation of methodologies and empirical strategies offers a preliminary insight into how challenging empirical and conceptual analyses of creative cognition can be. And indeed, “assessment has been a vexing problem for creativity researchers over the decades, in part because creativity research aspires to observe and measure things that are atypical, novel, innovative, and unusual, be they products, ideas, or people” (Silvia et al., 2012). Despite this difficulty, sometimes referred to as the “criterion problem” (Brown, 1989), the field has produced an array of analytical tools that permeate research and theory, often integrating quantitative and qualitative methodologies (e.g., Gaggioli et al., 2020). And while this combination has been more often adopted in contexts of individual creative production, it has also been applied to social contexts, where panels of experts are asked to judge various products and their creative properties (Amabile, 1982).

Along these lines, a number of studies concerned with the creative aspects of art and education have focused on how expressive qualities such as the experience of flow (e.g., Łuczniak et al., 2021) and compositional activity (e.g., Biasutti, 2015; 2018) can play out at the level of single agents as well as groups (see also Daikoku et al., 2021; Deliege & Wiggins 2006; Reybrouck, 2006; Wiggins, 2016 for more general approaches to musical creativity and inspiration). So, while classifications of individual and multi-agent creativities are expressed by different vocabularies and theoretical assumptions, they share similar empirical approaches and objects of investigation. The shift of unit of analysis, in other words, does not translate into new heuristics. This might give the impression that the separation between solo and collective creativity relies on a somewhat artificial move.² That is, although conceptually useful, one may question whether this distinction genuinely helps improve our understanding of creative cognition. Recent research and theory have indeed moved beyond this dichotomy, exploring the overlap between individual and collective creativity looking at both products and processes (Schiavio & Benedek, 2020).

The present work, as anticipated at the start of this section, is in direct alignment with this goal, as it focusses on assessing creativity when individuals generate a new artistic outcome tailored explicitly for themselves while also considering scenarios where they envision another person who could potentially derive value from it. As a result, this research explores a “social” scenario while maintaining its “individual” essence concerning the creative task (see Høffding & Satne, 2021; Schiavio et al., 2022a; 2022b). Also, it offers an in-depth examination of the qualities of creative outputs using participant self-assessments, expert ratings, and algorithmic evaluations. This

multifaceted approach bridges the study of processes and products, as well as individuality and collectivity, aiming to engage with the complexity of creative cognition in song-writing, specifically lyric generation. We have chosen this setting for our investigation as throughout history many individuals with diverse musical backgrounds have demonstrated the ability to blend instrumental music with *original* and *appropriate* lyrical content, resulting in a tapestry of narratives interwoven into a cohesive artistic whole.

Creative Metacognition

This focus on writing lyrics allows us to delve into the interplay between creativity and metacognition, illuminating the role of reflective and regulatory cognitive processes in the context of artistic production. Metacognition has been a recurring theme in various branches of psychology, with a primary focus on learning, memory, and educational perspectives (Norman et al., 2019). However, as noted by Lebuda and Benedek (2023), there is now a growing interest in exploring metacognition’s role in more complex cognitive activities such as reasoning, decision making, and problem solving (see e.g., Jia et al., 2019; Rivas et al., 2022). That said, conceptualizing metacognition remains a challenge, with ongoing debates around its structure, processes, and definition. Again following Lebuda and Benedek (2023), we observe that one of the most established proposals on the structure of metacognition involves two main components: metacognitive *knowledge* and metacognitive *experience*. The former contains explicit information related to tasks and strategies stored in long-term memory, while the latter reflects dynamic cognitive activities concerned with the monitoring and control of task performance.

Rhodes (2019) mentioned that a classic way to explore metacognition entails engaging participants in specific tasks, such as dart throwing, text reading, foreign language vocabulary learning, or general knowledge question answering. During these tasks, researchers would gather performance judgments from the participants. These judgments can be obtained through two methods: prospectively, before the criterion task, where participants assess their likelihood of success (e.g., “What is the probability of hitting the target?”), or retrospectively, after completing the criterion task, where participants evaluate the probability of their response being correct (e.g., “What is the likelihood that my answer to this question is correct?”). Through the juxtaposition of these assessments against the actual outcomes of the criterion task, researchers can discern two complementary indicators, one focused on the average distance to criterion measures and the other on correlations with criterion measures. An additional avenue for comparison emerges when participants’ appraisals are contrasted with those of a panel of experts (also known as Consensual Assessment Technique, see Amabile, 1982), thereby facilitating an analysis of both internal and external insight – a method that has worked efficiently in

investigations into performance creativity in sport (Gesbert et al., 2022).

However, in domains such as music, the establishment of unequivocal criteria remains elusive. This is particularly evident in the scrutiny of lyrics, where recent advancements in computational assessment, especially the *divergent semantic integration* (DSI) approach by Johnson et al. (2023), may offer novel avenues for analysis. The DSI approach evaluates how frequently words in a text co-occur next to each other based on a large corpus of text (e.g., all Wikipedia entries). Each word is represented in a multidimensional vector space, where frequently co-occurring words are situated closer to each other. To calculate the DSI score, the average of the distances between each word of a text is computed. Unusual combinations of words are reflected in a higher DSI score, which suggests a more creative text (Johnson et al., 2023). By employing DSI scores extracted from computational models, we could aspire to unveil more potentially “objective” benchmarks for gauging creativity, comparing them with self-assessments and external judgments. This convergence of heterogeneous viewpoints could in turn illuminate the cognitive processes underlying the creation of lyrics. Given the multifaceted significance of lyrics, encompassing social, artistic, emotional, and cognitive dimensions, such research not only enriches the realm of music but also affords insights into the broader panorama of human creativity.

Rationale for the Study

The present exploratory research builds on, and expands, the insights presented in the previous sections aiming (i) to increase our comprehension of both individual and collective creativity across the spectrum of processes and outcomes, and (ii) to examine how metacognitive insights can exert an influence on the dynamics of creative expression. To address this double and deliberately broad objective, the present study reports on an original empirical investigation that delves into the nuanced impact of various self-assessment metrics on novice and expert musicians while also examining correlations between these self-evaluations, computational creativity measures (DSI scores), and external judgments. We devised an experiment wherein participants were invited to generate novel song lyrics under two conditions: (a) envisioning themselves as the performers of the song, or (b) imagining someone else as the performer. After completing the lyric-creation task, participants engaged in reflective self-assessment concerning their emotional state (e.g., their enjoyment of the task), and behavior (e.g., if they took notes when writing lyrics during the trial), alongside providing self-rated scores for the creativity of their lyrics. Complementing this self-report data, we then enlisted the expertise of a number of songwriters to evaluate the creativity of the participants’ lyrics, and we also employed a computational assessment using DSI scores.

This experimental framework serves a twofold purpose. First, it enables us to examine the factors underlying participants’ self-perceived creativity in multiple ways. These factors could relate to the metacognitive dimension that dynamically guides the interaction between creative evaluation and production during the task. Additionally, it could help to explore whether altering the *social orientation* of the task influences changes in these factors. Ultimately, such work may lead to insights into how individuals incorporate social expectations and judgments into their metacognitive evaluation processes. Second, by juxtaposing computational and expert panel ratings with participants’ self-ratings, we can determine whether the factors influencing self-perceived creativity align with those shaping external evaluations. This comparison can illuminate potential congruence or divergence between these assessments. It should also be noted that our study encompasses both individuals with musical expertise and novices to capture any variances influenced by their level of proficiency. This deliberate inclusion acknowledges the importance of expertise in potential outcome disparities.

Methods

Participants

In the first phase of the study, participants were invited to write and self-assess the lyrics they wrote for two short instrumental pieces of music under two conditions. To do this, we initially recruited 95 participants via a dedicated platform (Prolific, www.prolific.co), social media, and large-group emails. We aimed for a sample size of a maximum of 100 participants to allow for a thorough evaluation of the lyrics’ creativity by our expert raters. Thirty-two participants were eventually removed from our data set as they had not followed the instructions properly. Excluded participants included those who uploaded PDF files without any lyrics (either intentionally or because they forgot to save the file properly), individuals who typed random letters instead of creating lyrics, or participants who typed their lyrics into a single text-field. It should be noted that such high exclusion rates are not uncommon in online experiments (Thomas & Clifford, 2017). Participants in online experiments might aim to complete experiments as fast as possible to increase their earnings, leading to missing and low-quality responses. The final data set consisted of 63 participants (23 men, 39 women, 1 non-binary; age: $M=29.59$, $SD=11.41$, $Range=18-68$). These were divided into two groups: *musical experts* ($n=35$; 13 men, 21 women, 1 non-binary; age: $M=27.06$, $SD=11.05$, $Range=18-68$) and *musical novices* ($n=28$; 10 men, 18 women; age: $M=32.75$, $SD=11.26$, $Range=20-62$). Participants were considered musical experts based on all the following criteria: (i) currently playing a musical instrument (which could include voice); (ii) having at least three years of formal training on a musical instrument (including voice); (iii) engaging in

regular, daily practice of a musical instrument (including voice) for at least three years.

For the second phase of the study, in which we evaluated the lyrics created by our participants, we recruited six *expert raters* (3 men, 3 women; age: $M=30.67$, $SD=11.54$, $Range=21-49$). They had at least 5 years of experience in songwriting ($M=12$, $SD=9.04$) and a minimum of 7 years of experience in performing music ($M=15.17$, $SD=5.19$). Expert raters individually rated all the lyrics via an online interface on a six-item questionnaire (see Table A3). The inter-rater agreement across all rated lyrics amounted to at least 70%³ for each rating item, except for the question asking to assess how well the lyrics go with the music (i.e., can be sung efficiently). For this question, the inter-rater agreement could not be estimated as the fitted linear mixed model was singular, that is, no unique solution could be found for the model's parameters. Inter-rater agreement was assessed by calculating the intraclass correlation coefficient (ICC) as described by Koo and Li (2016). ICC estimates and 95% confidence intervals were calculated using the R package "psych" (Revelle, 2025) based on a mean rating ($k=6$), absolute agreement, two-way mixed effect model. All participants gave informed consent, and the study was approved by the ethics committee of the University of Graz.

Tasks and Materials

Stimuli for the study included an instructional video and two short instrumental musical excerpts. As described below, the latter were administered through videos providing a visual representation of the melodic lines. The instructional video had a duration of 4 min and offered a debriefing of the various tasks of the experiment (how to fill out the forms, where to write the lyrics, etc.), showing concrete examples. The two short musical excerpts were composed by the first author for the purpose of the study (available upon request). The first excerpt ("song1") had a duration of 41 s; it has a melodic line played by the flute, accompanied by piano, strings, and drums. The tempo is moderate (quarter = 71) and is in D minor with a modulation into C major towards the end. The second stimulus ("song2") has a duration of 53 s and also has its melodic line played by the flute, accompanied by drums and other instruments. The "character" of this latter piece is different from "song1" in that it has a livelier course (quarter = 98), and a major key (it is in D major with a modulation into B minor towards the end). In both cases, participants were invited to write their own lyrics thinking of the melodic line of the flute as a *vocal line*, on which the created lyrics can be sung. The musical excerpts were created via software MuseScore (www.musescore.org, version 2) as MIDI files. They were then extracted as .wav files and presented to participants along with a video offering a piano-roll visualization of the melodic line (see the videos provided in the repository in the data availability statement). The piano-roll visualization was generated by the software Audacity

(Audacity, audacityteam.org). In a piano-roll visualization, notes are displayed as horizontal bars on a grid, where the vertical axis represents pitch, and the horizontal axis represents time. Melody and rhythm of a musical piece are, hence, clearly depicted. The videos of the songs, including the piano roll visualization, were particularly useful for non-musicians as they did not show notation; instead of the notes, they showed a succession of a series of boxes at different heights representing the notes of the melodic line (see Fig. 1). We chose this option because, as outlined below in more detail, in the subsequent phase of the experiment participants were invited to download a PDF file displaying the same visual representation of the songs (including boxes) they saw on video and write their lyrics directly into these boxes. By doing so, we wanted to ensure that participants had a clear sense of the different notes that made up the vocal melodies, so that they would be aware that words must often be broken down into syllables to be sung properly.

Questionnaires: The following questionnaires were administered to the participants before generating lyrics: 1) General demographic and background questions (see Table A1); 2) the Assessment of Everyday Creativity Across Nine Domains "AEC" (Benedek et al., 2019); 3) Top 3 Creative Achievements "T3CA" similar to Diedrich and colleagues (2018) (see also Ceh et al., 2022); 4) Big Five Inventory "Openness to experience" subscale "BFI-O" (John & Srivastava, 1999); 5) Interpersonal Reactivity Index Perspective-taking "IRI-P" and Empathetic concern subscale "IRI-E" (Davis, 1980); 6) the Goldsmiths Musical Sophistication Index, items 32–39 "GMSI" (Müllensiefen et al., 2014). After generating lyrics, participants completed another questionnaire to assess their behavior and states during the task. This included questions on, for example, how difficult the participants found the task, whether they took notes, and how often they listened to the songs. The questionnaire was adapted for each condition (see Table A2). A six-item questionnaire (see Table A3) based on the main characteristics of creativity as outlined by Runco and Jager (2012) was given to the expert raters so they could assess each set of lyrics. This questionnaire included six Likert items with a seven-point response range ranging from "Strongly disagree" to "Strongly agree." Items asked for appropriateness ("how well the lyrics go with the music"), creativity, originality, surprisingness, and likeability.

Procedure

Participants took part in the study remotely, with their own computer. They either received a link via email or accessed the experiment portal through the Prolific participant recruitment platform (Prolific, prolific.co) directly. The study was implemented using the *jsPsych* JavaScript framework (de Leeuw, 2015); it was administered online using Pavlovia (www.pavlovia.org) and was carried out in two phases. While the first phase involved *musical experts*

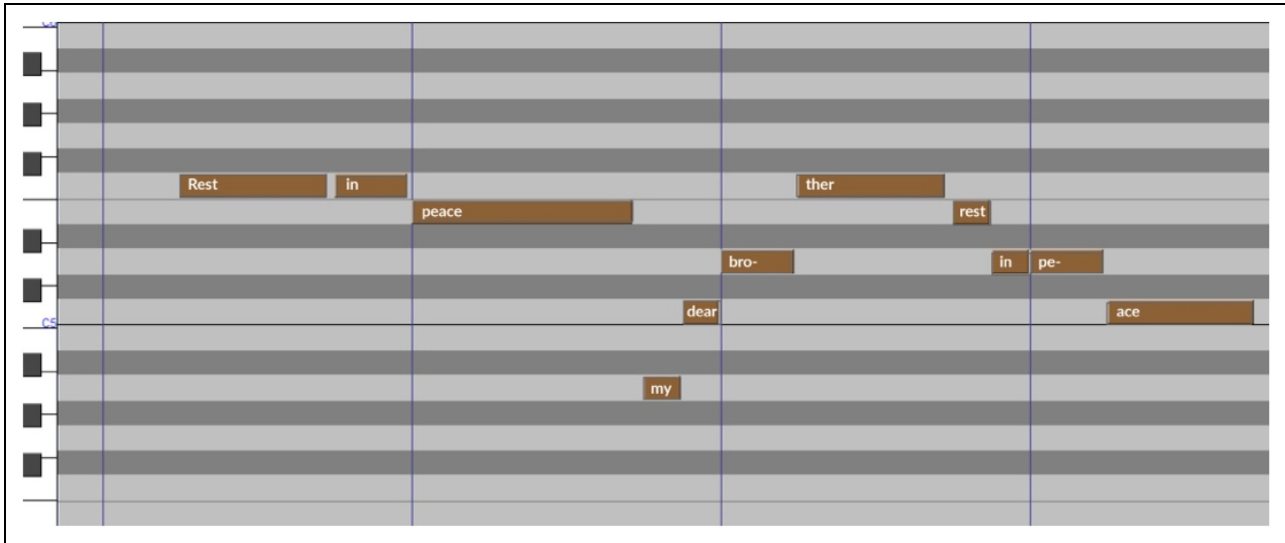


Figure 1. Part of the lyrics and the piano-roll visualization of song1 as submitted by one participant.

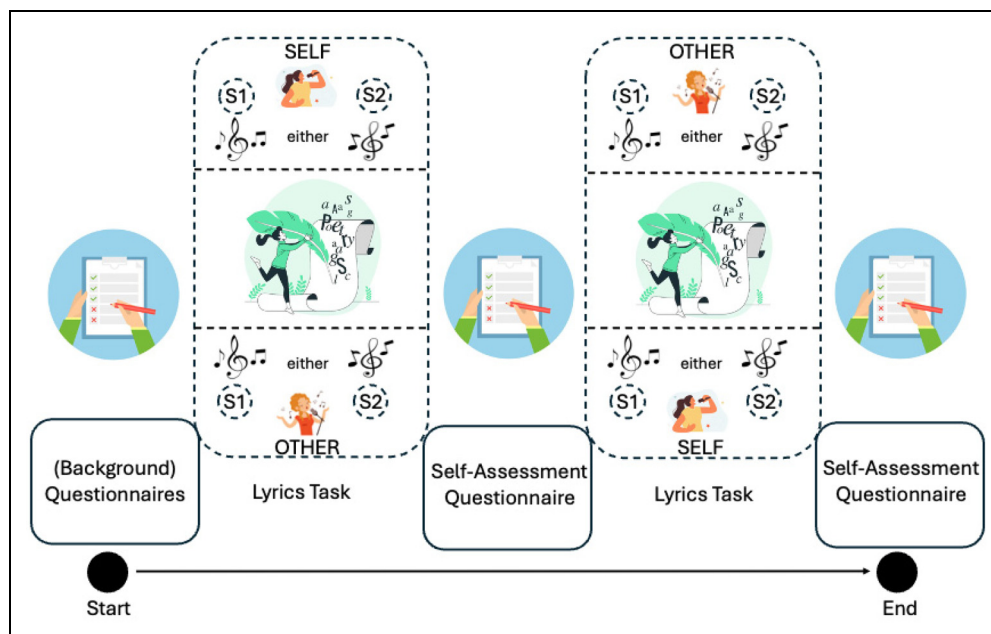


Figure 2. Experimental procedure of the first phase of the study. Participants started with a background questionnaire and afterwards ran through two conditions (fully counterbalanced) in which they were asked to create lyrics for two musical excerpts, or songs (i.e., S1 and S2). They had to either write a song for themselves, imagining they would be the singer (SELF), or for someone else, imagining any other person being the singer (OTHER). Each condition featured a different song.

and *musical novices*, the second one only involved *expert raters*.

Phase 1: At the beginning of the experiment, musical experts and musical novices completed a general background questionnaire as well as the AEC, T3CA, BFI-O, IRI-P, IRI-E and GMSI questionnaires. After this, they watched a 4-min instructional video introducing them to the tasks of the experiment. Here, participants wrote lyrics for two short musical excerpts (“song1” and “song2”) in two separate trials. The order of conditions

and songs was fully counterbalanced, giving rise to four counterbalanced blocks (2 conditions x 2 songs). In the condition SELF, participants were invited to write the lyrics *for themselves* (imaging they would be the singer); in the condition OTHER they instead wrote the lyrics as if *someone else* (any other person they could think of but themselves) were to sing them. As mentioned above, participants were initially presented with a video featuring the series of boxes and a piano-roll visualization of the melody line to which they should write the lyrics. During the trial they

Table 1. Linear mixed model (M.1) to evaluate the association between participants' creativity self-assessments of their lyrics and their reported behaviors and states during the lyric writing task.

(M.1)

creativity self-assessment ~ *musical_expertise* * (*condition* + *notes* + *enjoyment* + *time* + *difficulty*) + (1|*participant*)

Predictors	β	95% CI	p
(Intercept)	−0.01	[−0.24, 0.21]	.901
musical_expertise	0.08	[−0.14, 0.30]	.476
condition	< 0.01	[−0.13, 0.14]	.895
notes	−0.06	[−0.27, 0.15]	.575
enjoyment	0.29	[0.10, 0.47]	.002
time	−0.04	[−0.25, 0.17]	.714
difficulty	−0.01	[−0.20, 0.18]	.916
musical_expertise (Experts) * condition (OTHER)	−0.12	[−0.25, 0.01]	.080
musical_expertise (Experts) * notes (Yes)	0.16	[−0.05, 0.38]	.132
musical_expertise (Experts) * enjoyment	−0.06	[−0.25, 0.12]	.497
musical_expertise (Experts) * time	< 0.01	[−0.22, 0.20]	.936
musical_expertise (Experts) * difficulty	−0.07	[−0.26, 0.12]	.466
Random Effects			
σ^2	0.73		
τ_{00} participant	0.62		
ICC	0.41		
N participant	63		
Observations	126		
Marginal R_m^2 / conditional R_c^2	0.169 / 0.514		

σ^2 (Residual variance); τ_{00} (Between-group-variance); ICC (Intraclass correlation coefficient); Marginal R_m^2 (Explained variance by fixed effects); Conditional R_c^2 (Explained variance by fixed and random effects).

could watch the video as many times as they wished to. They were instructed that when they felt ready to write the lyrics, they could stop watching the video and download a PDF file with the same piano-roll visualization of the melody of the video. Within the PDF file, each line (representing a note) in the piano-roll visualization included a text field where participants could enter lyrics. Upon completion, participants were instructed to save the PDF file. Each writing trial was ensued by a post-questionnaire adapted for each condition. Finally, participants were asked to upload the two PDF files with their lyrics via a simple file upload dialog. The procedure is illustrated in Fig. 2.

Phase 2: All lyrical content generated by both musical experts and musical novices was assessed by six expert songwriters. The raters evaluated all the lyrics for one song in a continuous block before resuming with the lyrics of the other song. The order of songs was randomized. During the rating procedure, the experts could always watch a video featuring the song and a visualization of the melody. Experts evaluated the lyrics based on the piano-roll visualization (see again Fig. 1), allowing them to examine the placement of syllables across notes, to also assess their effectiveness.

Data Analysis

Our analysis was carried out using R (R Core Team, 2025). For the computation of linear mixed models, we used the

package “lme4” (Bates et al., 2015) and applied a sum contrast coding scheme to all categorical variables. Due to the sum contrast scheme, the linear regression tested the difference between the mean across a categorical variable and each value of this variable. In the case of a dichotomous variable, the regression simply tested the difference between the two values of this variable. The marginal R_m^2 and conditional R_c^2 values are calculated by the “sjPlot” (Lüdtke, 2024) package using the method proposed by Nakagawa and colleagues (2017). All reported regression coefficients are standardized. We decided to use linear mixed models in our analysis as we had a twofold aim: to inspect differences in creativity regarding participants' behavior and state during the task and to estimate whether these factors (measurements of behaviors and states) predict participants' creativity in writing lyrics. Other measurements (e.g., everyday creativity (AEC) or openness to experience (BFI-O)), were not included in the final analysis as we aimed to explore solely the influence of the participants' experience during the task. Yet, the experimental design made it necessary to include the condition and the musical expertise (expert/novice) as confounding factors in our analysis. By using linear mixed models and carrying out post-hoc comparisons we could complete both aims using a single model. The linear mixed models and their coefficients are reported in detail in Tables 1–3, while post-hoc comparisons are reported within the running text. The analysis focused on three different levels of creativity assessments: 1) *Self-assessment* of the creativity of

Table 2. Linear mixed model (M.2) to evaluate the association between the expert creativity ratings of the lyrics and the participants' reported behaviors and states during the lyric writing task.

(M.2)

$$\text{expert creativity rating} \sim \text{musical_expertise} * (\text{condition} + \text{notes} + \text{enjoyment} + \text{time} + \text{difficulty}) + (1|\text{participant}) + (1|\text{rater})$$

Predictors	β	95% CI	p
(Intercept)	−0.02	[−0.40, 0.36]	.908
musical_expertise	0.12	[−0.01, 0.25]	.063
condition	< 0.01	[−0.05, 0.07]	.753
notes	−0.06	[−0.17, 0.05]	.299
enjoyment	−0.02	[−0.12, 0.07]	.597
time	0.12	[0.01, 0.22]	.033
difficulty	0.03	[−0.07, 0.12]	.575
musical_expertise (Experts) * condition (OTHER)	> −0.01	[−0.06, 0.06]	.889
musical_expertise (Experts) * notes (Yes)	0.07	[−0.04, 0.18]	.223
musical_expertise (Experts) * enjoyment	−0.03	[−0.12, 0.06]	.501
musical_expertise (Experts) * time	0.06	[−0.05, 0.17]	.283
musical_expertise (Experts) * difficulty	−0.13	[−0.23, −0.03]	.009
Random Effects			
σ^2	0.80		
τ_{00} participants	0.41		
τ_{00} raters	0.45		
ICC	0.36		
N participants	63		
N raters	6		
Observations	756		
Marginal R_m^2 / conditional R_c^2	0.046 / 0.394		

σ^2 (Residual variance); τ_{00} (Between-group-variance); ICC (Intraclass correlation coefficient); Marginal R_m^2 (Explained variance by fixed effects); Conditional R_c^2 (Explained variance by fixed and random effects).

Table 3. Linear mixed model (M.3) to evaluate the association between the algorithmic assessments of the lyrics (divergent semantic integration; DSI) and the participants' reported behaviors and states during the lyric writing task.

(M.3)

$$\text{DSI} \sim \text{musical_expertise} * (\text{condition} + \text{notes} + \text{enjoyment} + \text{time} + \text{difficulty}) + (1|\text{participant})$$

Predictors	β	95% CI	p
(Intercept)	−0.06	[−0.28, 0.15]	.545
musical_expertise	0.24	[0.03, 0.46]	.024
condition	> −0.01	[−0.14, 0.13]	.947
notes	−0.06	[−0.26, 0.14]	.566
enjoyment	0.04	[−0.14, 0.22]	.643
time	0.14	[−0.06, 0.34]	.173
difficulty	0.20	[0.02, 0.39]	.031
musical_expertise (Experts) * condition (OTHER)	−0.02	[−0.16, 0.11]	.713
musical_expertise (Experts) * notes (Yes)	0.22	[0.01, 0.42]	.037
musical_expertise (Experts) * enjoyment	−0.08	[−0.26, 0.09]	.358
musical_expertise (Experts) * time	0.02	[−0.17, 0.22]	.814
musical_expertise (Experts) * difficulty	−0.20	[−0.38, −0.01]	.036
Random Effects			
σ^2	0.72		
τ_{00} participants	0.57		
ICC	0.39		
N participants	63		
Observations	126		
Marginal R_m^2 / conditional R_c^2	0.208 / 0.515		

σ^2 (Residual variance); τ_{00} (Between-group-variance); ICC (Intraclass correlation coefficient); Marginal R_m^2 (Explained variance by fixed effects); Conditional R_c^2 (Explained variance by fixed and random effects).

the lyrics, that is, how creative the participants rated their own lyrics for each song; 2) *Expert ratings* of the lyrics' creativity, that is, how creative our expert songwriters rated the lyrics of each participant; 3) *Algorithmic assessment* of the creativity of the lyrics, that is, a computational assessment of the lyrics' creative features using distributional semantic modelling (Johnson et al., 2023).

Predicting creativity self-assessments: We used the participants' single-item self-assessment of the creativity of their own lyrics as the outcome variable of the linear mixed model (M.1). The model also included a varying intercept for participant. First, we included the categorical variable "condition" as a predictor to test the effect of the social context (either writing the song for themselves or for someone else to perform). Furthermore, we were particularly interested if the participants' behavior and state during the songwriting task would predict their creativity ratings. Thus, we included the following predictors: 1) "enjoyment" / How much they enjoyed the music. 2) "time" / How long they reported it took them to finish the task. 3) "difficulty" / How difficult they found the task. 4) "notes" (categorical variable) / If they took notes during the task. As we had two groups of different expertise, we also added the binary categorical variable of "musical_expertise" including the interaction with all the other factors as predictors. To assess differences in creativity self-assessments dependent on factors such as "musical_expertise" or "condition" we performed post-hoc comparisons using the "emmeans" package (Lenth, 2025).

Expert ratings: We then investigated whether the evaluations of creativity offered by our experts were affected by condition, musical expertise, or related to the behaviors and states of our participants during the writing task. The final expert creativity ratings were computed by taking the mean (on the level of each individual rating) of the two items assessing creativity and originality, as they are highly correlated, with a Cronbach Alpha value of $\alpha = .89$, 95% CI [.87, .9]. The average expert creativity rating was then used as the outcome variable in our linear mixed model (M.2). The same predictors as in model (M.1) were chosen for model (M.2). In addition to the varying intercept for participant we also included a varying intercept for rater.

Algorithmic assessment: Finally, we examined whether assessing the creativity of the lyrics using a computational approach leads to results comparable to the evaluations of the expert raters. Recently, the computational method of distributional semantic modelling has been used to rate the creativity of short stories by calculating a divergent semantic integration (DSI) score for each story (Johnson et al., 2023). The DSI score is estimated by calculating the distance between each pair of words within a response and averaging over these distances. The semantic distance roughly reflects how often two words can be found within the same context. These distances are represented within a semantic model that has been constructed from a large database of texts such as Wikipedia (www.wikipedia.org). In our study, we employed the bidirectional encoder transformer

"BERT" semantic model used by Johnson and colleagues (2023). BERT language models can capture nuanced relationships between words in a text as context-dependent meanings (e.g., whether the word "rose" refers to the flower or the past simple form of "to rise"). A BERT model also assesses and weighs how important each word in a sentence is. These characteristics likely contribute to the model's superior performance in predicting linguistic creativity, reaching a correlation of around .76 with human expert ratings. Our aim was to explore whether a BERT model would perform similarly well when assessing the creativity of lyrics for music. DSI scores were calculated from the script by Johnson and co-workers (2022) using the latest (as of August 29, 2022) pre-trained BERT-Large model from the Google research collection. We used the same predictors and random effect structure as in model (M.1) to estimate model (M.3), which included the DSI score as dependent variable.

Results

Self-Assessments

Our analysis revealed that participants' behavior and state during the songwriting task are highly predictive of their creativity self-assessment ratings of their lyrics. As illustrated in Table 1, the independent variables of our model (M.1) explain a substantial proportion of $R_m^2 = .17$ of the variation in participants' creativity ratings. Inspecting the factors at play in detail, we found a significant ($\beta = .29$, 95% CI [.10, .47]) main effect of "enjoyment." Hence, people who enjoyed the music more also rated their lyrics as more creative.

Expert Ratings

We then tried to predict the creativity of the lyrics, as judged by our expert raters (M.2), considering the same predictors we used in model (M.1). As illustrated in Table 2, the independent variables explained only $R_m^2 = .05$ of the variation of the dependent variable in model (M.1). We found a main effect of "time," indicating that the time it took participants to complete the task was significantly ($\beta = .12$, 95% CI [.01, .22]) correlated with higher expert creativity ratings. Moreover, our results indicate that musical experts show a significantly weaker correlation ($\beta = -.13$, 95% CI [-.23, -.03]) between the self-rated difficulty of the task and the experts' creativity ratings when compared to musical novices. A post-hoc comparison revealed that for novices, self-rated difficulty is significantly correlated with experts' creativity ratings (Novices: $M_{\text{difficulty}} = 0.13$, 95% CI [0.01, 0.24]), while for musical experts the correlation was not significant (Experts: $M_{\text{difficulty}} = -0.08$, 95% CI [-0.18, 0.02]).

Algorithmic Assessment

Finally, we aimed to explore the performance of a BERT model in rating the creativity of lyrics. Similar to the

work by Johnson et al. (2023), we calculated a Pearson correlation between the average expert ratings and the DSI score for each of our participants' lyrics. We found the correlation between these variables to reach $r = .44$ (95% CI [.29, .57]). Out of interest we also calculated the correlation between the DSI score and participants' self-assessment ratings of the creativity of their lyrics, which only amounted to $r = .25$ (95% CI [.08, .41]). The correlation between participants' self-assessment ratings and the experts' creativity ratings was not significant ($r = .13$, 95% CI [−.05, .30]).

We computed the linear mixed model (M.3) to estimate to what degree the participants' reported states and behaviors predicted the DSI scores of their lyrics. In model (M.3) the independent variables were found to explain an $R_m^2 = .21$ of the variation of the DSI scores. We found a significant main effect of musical expertise ($\beta = .24$, 95% CI [.03, .46]), with a post-hoc comparison showing that musical experts exhibited significantly higher DSI scores than musical novices (Experts – Novices: $M = 0.49$, $t(61.7) = 2.286$, $p = .026$). Furthermore, our analysis revealed that self-rated difficulty correlated significantly ($\beta = .20$, 95% CI [.02, .39]) with the DSI score. We also found a significant interaction between musical expertise and self-rated difficulty ($\beta = -.20$, 95% CI [−.38, −.01]). Post-hoc comparisons show that the correlation between self-rated difficulty and DSI scores is only significantly different from zero for novices (Novices: $M_{difficulty} = 0.22$, 95% CI [0.07, 0.38]) but not for experts (Experts: $M_{difficulty} = 0.003$, 95% CI [−0.14, 0.15]). We performed additional post-hoc comparisons to inspect the significant interaction ($\beta = .22$, 95% CI [.01, .42]) between musical expertise and taking notes in detail. These analyses showed that among the participants who did not take notes, experts performed significantly better (Experts – Novices: $M_{notes(No)} = 0.92$, $t(69.9) = 3.764$, $p < .001$) than novices. Such a difference between musical experts and novices could not be found for the participants taking notes (Experts – Novices: $M_{notes(Yes)} = 0.05$, $t(82.4) = 0.158$, $p = .875$).

(Figure 3).

Discussion and Conclusion

This study has examined the creative output and experience of musical novices and experts in a song lyric writing task. The central aim was to investigate the factors underlying perceptions of creativity, and this was achieved by comparing self-, expert, and computational assessments of song lyric creativity. Our intention was to study participants' experiences of the writing task, exploring differences relating to demands of either writing lyrics as though for oneself to sing, or for an imagined other.

A main finding was the positive relationship between participants' enjoyment of the task and their positive rating of the self-assessed creativity of their song lyrics in model (M.1). This is consistent with research on creative metacognitive feelings (Puente-Diaz & Cavazos-Arroyo, 2018) and creative self-efficacy, defined as belief in one's

own creative ability (Puente-Diaz, 2015), and more generally with enjoyment being the strongest motive for creative behavior (Benedek et al., 2020). For example, Puente-Diaz and Cavazos-Arroyo (2018) found evidence that metacognitive feelings such as perceived ease of generating ideas positively influenced task enjoyment via increased creative self-efficacy. It would be interesting to explore creative self-efficacy as a mediator between task enjoyment and confidence in the creative outcome in musical tasks. Also, enjoyment predicted self-assessed creativity but not expert- or algorithmically-scored creativity, highlighting the self-relevant yet partly subjective nature of creative engagement. Of interest here is the possible interrelationship between metacognitive decisions such as whether to engage in notetaking as a tool to support creativity (Lebuda & Benedek, 2023) and creative beliefs such as the extent to which the nature of creativity is, or should be, intuitive or spontaneous. For example, expert participants may have felt as though more spontaneous processes (without the need to make notes) correspond to more intuitive and inspired creative practice. Conversely, musical novices may be less inclined to trust their intuition, having potentially lower degrees of creative self-efficacy; this could mean that they perceived notetaking to have facilitated their creative work, serving as a scaffolding to their outputs.

A further distinction between musical experts and novices that emerges from our findings is the lower correlation between self-perceived task difficulty and creative success for experts than for novices, as found in model (M.2). Although the musical expert participants were not recruited as song writers, their broader musical knowledge of melodic and rhythmic structures may have improved their confidence in setting lyrics to the stimuli and supported their ability to evaluate their resulting creative efforts. In contrast, novice participants without these foundational levels of domain-relevant knowledge may be more influenced by metacognitive factors such as their assessment of task difficulty when reflecting on their creative outcomes. Interestingly, in none of our analyses did we find musical expertise to interact with the social perspective that participants were asked to adopt when writing their song lyrics, that is, as though writing for themselves to sing or as though someone else would sing them. This contrasts with work by Lu and colleagues (2023) that suggests that experts may inhibit social information in creative tasks to focus on more self-centered processes (Lu et al., 2023). Instead, we found that experts and novices did not show any difference in self-ratings of creativity when writing lyrics for themselves or someone else. Of interest here is the perspectival framework outlined by Glăveanu (2015) which emphasizes that creativity is a sociocultural process, involving a dialogue between self/other perspective-taking and self-reflection. He argues that “creative action emerges [...] when we are able to ‘move’ (imaginatively and/or practically)” (p. 171) between perspectives. This ability is arguably equally developed in both novices and experts, suggesting that musical creativity can be nurtured from the

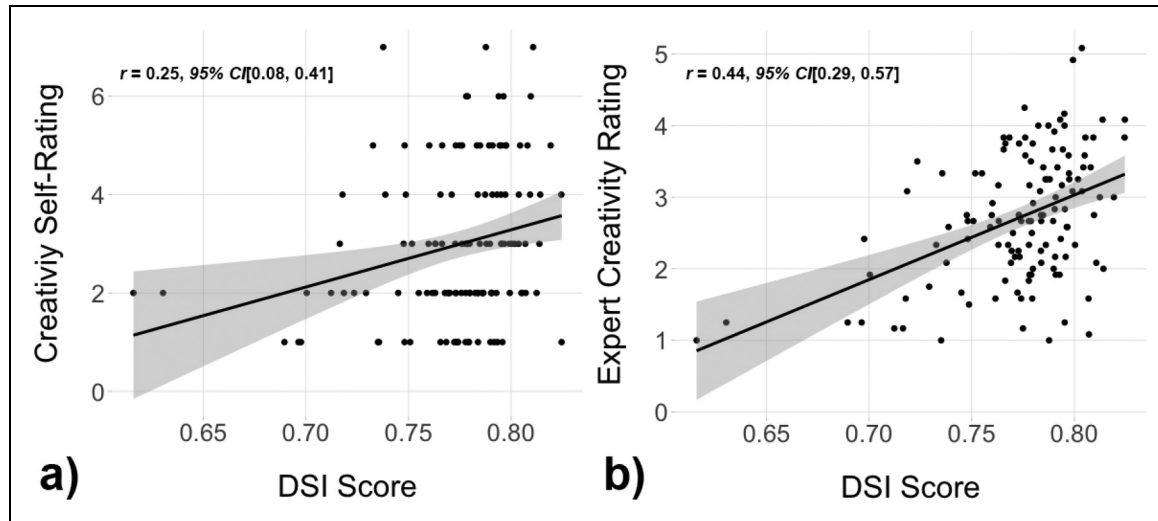


Figure 3. Correlation between the DSI (divergent semantic integration) scores as computed using the BERT model and a) participants' creativity self-ratings and b) the expert creativity ratings respectively.

very beginning of one's musical journey and arguably does not fully depend on musical skills and knowledge (see Schiavio & Nijs, 2022).

Although the role of musical expertise was not relevant for differentiating between groups in self- and expert-ratings, our computational analysis in model (M.3) indicates that musical experts exhibit higher DSI scores compared to non-experts. This observation is particularly intriguing as DSI scores pertain exclusively to the verbal domain, whereas raters also considered how lyrics were embedded within the musical material. This raises the question of whether it is verbal aptitude – such as the ability to manipulate language and compose lyrics – rather than musical proficiency that may account for the observed differences. An alternative explanation could be that effective word setting, such as writing lyrics, inherently involves a sensitivity to various musical elements as rhythm, timbre, contour, and others. Musical experts might be more aware of these musical elements (Carey et al., 2015). These musical considerations might lead lyricists with musical expertise to select less common word conjunctions that better fulfil the musical demands of word setting, which could, in turn, be reflected in the DSI scores.

The results of the computational analysis in model (M.3) reveal that novices who find the task more challenging tend to produce lyrics that are rated as more creative by our algorithm. In contrast, experts exhibit a significantly lower correlation between computational creativity ratings and their output, though this difference ultimately fails to reach statistical significance. Lastly, it is not surprising that experts who did not take notes performed significantly better than novices who also refrained from notetaking (see the results in model (M.3)). Our analyses show how the different assessments of creativity complement each other by picking up different aspects of the creative process: Task enjoyment is linked to increased self-ratings; time spent

on the task is positively correlated with expert ratings; and for novices, task effort is associated with higher expert ratings and DSI scores. Overall, there is a greater alignment between self-perceived creativity and participants' subjective experiences of the task in model (M.1) than with external judgements of lyrical creativity in model (M.2). Even more interestingly, comparing model (M.3) to the other models shows that participants' subjective experience of the task is better in predicting DSI scores than self-ratings and especially expert ratings of creativity. Yet the assessments of creative success from the expert songwriters are informative (see model (M.2) for the analysis). For example, it is interesting to note that more objective task behaviors, such as time taken to write the lyrics, positively predicted the lyrical creativity judgements of the external songwriters. Expert creativity ratings diverge in their alignment with more subjective measures of perceived task difficulty experienced by the novice and expert musicians: Lower creativity ratings were associated with higher levels of perceived task difficulty for the musical novices, where no such relationship was found for the musical experts. Future research may benefit from the insights afforded by employing the Consensual Assessment Technique (CAT) (Amabile, 1982). Here, a panel of experts is invited to judge the creativity of the musical (in this case lyrical) product, through a process of refining their collective judgement of what may constitute creative success. Qualitative data regarding what is subjectively important for song lyric raters can only enrich our understanding.

In the current study, we instead took a further step toward objectivity through the addition of the calculation of DSI scores. Our main finding was a moderately strong correlation with the expert raters' evaluations of lyrical creativity and substantially less agreement with participants' own assessment of creativity. As discussed, it seems as though participants' assessment of their creative success is strongly

influenced by their subjective experience of performing the task. A range of metacognitive factors can be expected to dynamically shape the experience of task success from an initial evaluation of the task characteristics and anticipation of one's own performance, through the on-task evaluation of success, to the retrospective evaluation of creative success (see Lebudá & Benedek, 2023).

To conclude, our study demonstrates how different measures of creativity reflect different aspects of individuals' experience in creative song lyric writing. An important finding is the relevance of metacognitive experience in the self-assessment of creativity, while participant experience is less related to expert assessments of creative performance. This finding again demonstrates that expert- and self-assessments of creativity should not be seen as exchangeable but rather as complementary measures reflecting different aspects (Kaufman, 2019). Interestingly, the computational evaluation of lyrics using DSI scores is associated with expert-ratings than with self-ratings. Future research may wish to further probe the parallels between the DSI scores of song lyrics and the expert songwriters' views of lyrical creativity. Setting words to music is a musical task that extends beyond integrating semantic divergent concepts (as assessed by the DSI scores), and there is scope to apply a wider range of computational methods to investigate different dimensions of melodic, harmonic, and rhythmic creativity (see Pearce, 2005).

Acknowledgments

This research was funded in whole, or in part, by the Austrian Science Fund (FWF) (10.55776/P32460). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Action Editor

Emily Payne, University of Leeds, School of Music.

Peer Review

Mark Reybrouck, KU Leuven, Department of Musicology.
Mihailo Antović, University of Niš, Faculty of Philosophy.

Contributorship

A.S. and A.K. were responsible for conceptualization, methodology, validation, resources, project administration, writing – original draft, and writing – review & editing. A.K. was responsible for software, formal analysis, investigation, data curation, and visualization. A.S. was responsible for supervision and funding acquisition. M.B. and F.B. were responsible for conceptualization and writing – review & editing.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval


The study was approved by the ethics committee of the University of Graz.


Funding


The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Austrian Science Fund, (grant number 10.55776/P32460).

ORCID iDs

Andrea Schiavio  <https://orcid.org/0000-0001-8109-9185>

Adrian Kempf  <https://orcid.org/0000-0002-0428-3090>

Freya Bailes  <https://orcid.org/0000-0003-2723-3579>

Mathias Benedek  <https://orcid.org/0000-0001-6258-4476>

Data Availability Statement

The video stimuli used in this study, along with the dataset utilized for the analysis, can be found in the following repository: https://osf.io/bmq2n/?view_only=c74b07d854424a11b3be4c9608929aa3

Notes

1. Celebrated lyricists like Lorenzo Da Ponte, who made pivotal contributions to Mozart's operas, and contemporary artists like Bernie Taupin and Hal David might exemplify the enduring importance of lyrics generation in music.
2. As Von Held (2012) observes, there have been previous attempts to reconcile individual and collective elements in multi-modal approaches (Amabile, 1996; Feldman, 1999).
3. A minimum inter-rater agreement of 70% is typical for creativity studies relying on expert creativity ratings (Barth & Stadtmann, 2021).

References

- Abraham, A. (2018). *The Neuroscience of Creativity*. Cambridge University Press.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43(5), 997. <https://doi.org/10.1037/0022-3514.43.5.997>
- Amabile, T. M. (1996). *Creativity in Context*. Westview Press.
- Antović, M., Küssner, M., Kempf, A., Omigie, D., Hashim, S., & Schiavio, A. (2023). A huge man is bursting out of a rock". bodies, motion, and creativity in verbal reports of musical connotation. *Journal of New Music Research*, 52(1), 73–86. <https://doi.org/10.1080/09298215.2024.2306406>
- Barth, P., & Stadtmann, G. (2021). Creativity assessment over time: Examining the reliability of CAT ratings. *The Journal of Creative Behavior*, 55(2), 396–409. <https://doi.org/10.1002/jocb.462>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benedek, M., Bruckdorfer, R., & Jauk, E. (2020). Motives for creativity: Exploring the what and why of everyday creativity. *J. Creat. Behav*, 54, 610–625. <https://doi.org/10.1002/jocb.396>

- Benedek, M., Christensen, A. P., Fink, A., & Beaty, R. E. (2019). Creativity assessment in neuroscience research. *Psychol. Aesthet. Creat. Arts*, 13, 218–226. <https://doi.org/10.1037/aca0000215>
- Biasutti, M. (2015). Creativity in virtual spaces: Communication modes employed during collaborative online music composition. *Think. Skills Creat.*, 17, 117–129. <https://doi.org/10.1016/j.tsc.2015.06.002>
- Biasutti, M. (2018). Strategies adopted during collaborative online music composition. *Int. J. Music Educ.*, 36, 473–490. <https://doi.org/10.1177/0255761417741520>
- Bissola, R., & Imperatori, B. (2011). Organizing individual and collective creativity: Flying in the face of creativity clichés. *Creativity and Innovation Management*, 20(2), 77–89. <https://doi.org/10.1111/j.1467-8691.2011.00597.x>
- Brown, R. T. (1989). Creativity: What are we to measure? In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 3–32). Plenum Press.
- Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, 137, 81–105. <https://doi.org/10.1016/j.cognition.2014.12.005>
- Ceh, S. M., Edelmann, C., Hofer, G., & Benedek, M. (2022). Assessing raters: What factors predict discernment in novice creativity raters? *Journal of Creative Behavior*, 56(1), 41–54. <https://doi.org/10.1002/jocb.515>
- Cook, N. (2018). *Music as creative practice*. Oxford University Press.
- Daikoku, T., Wiggins, G. A., & Nagai, Y. (2021). Statistical properties of musical creativity: Roles of hierarchy and uncertainty in statistical learning. *Front. Neurosci.*, 15, 640412. <https://doi.org/10.3389/fnins.2021.640412>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85–103.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Deliege, I., & Wiggins, G. A. (2006). *Musical creativity: Multidisciplinary research in theory and practice*. Psychology Press.
- Diedrich, J., Jauk, E., Silvia, P. J., Gredlein, J. M., Neubauer, A. C., & Benedek, M. (2018). Assessment of real-life creativity: The inventory of creative activities and achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts*, 12(3), 304–316. <https://doi.org/10.1037/aca0000137>
- Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Pers. Soc. Psychol. Rev.*, 2, 290–309. https://doi.org/10.1207/s15327957pspr0204_5
- Feldman, D. H. (1999). The development of creativity. In J. Sternberg (Ed.), *Handbook of creativity* (pp. 169–186). Cambridge University Press.
- Gaggioli, A., Mazzoni, E., Benvenuti, M., Galimberti, C., Bova, A., Brivio, E., et al. (2020). Networked flow in creative collaboration: A mixed method study. *Creativity Research Journal*, 32(1), 41–54. <https://doi.org/10.1080/10400419.2020.1712160>
- Gesbert, V., Hauw, D., Kempf, A., Blauth, D., & Schiavio, A. (2022). Creative togetherness. A joint-methods analysis of collaborative artistic performance. *Frontiers in Psychology*, 13, 835340. <https://doi.org/10.3389/fpsyg.2022.835340>
- Glăveanu, V. P. (2013). Rewriting the language of creativity: The five a's framework. *Rev. Gen. Psychol.*, 17, 69–81. <https://doi.org/10.1037/a0029528>
- Glăveanu, V. P. (2015). Creativity as a sociocultural act. *The Journal of Creative Behavior*, 49(3), 157–244. <https://doi.org/10.1002/jocb.94>
- Guildford, J. P. (1950). Creativity. *American Psychologist*, 5, 444–454. <https://doi.org/10.1037/h0063487>
- Hill, J. (2018). *Becoming creative: Insights from musicians in a diverse world*. Oxford University Press.
- Høffding, S., & Satne, G. (2021). Interactive expertise in solo and joint musical performance. *Synthese*, 198(Suppl 1), 427–445. <https://doi.org/10.1007/s11229-019-02339-x>
- Jia, X., Li, W., & Cao, L. (2019). The role of metacognitive components in creative thinking. *Frontiers in Psychology*, 10, 2404. <https://doi.org/10.3389/fpsyg.2019.02404>
- John, O. P., & Srivastava, S. (1999). The big-five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). Guilford Press.
- Johnson, D. R., Kaufman, J. C., Baker, B. S., Patterson, J. D., Barbot, B., Green, A. E., van Hell, J., Kennedy, E., Sullivan, G. F., Taylor, C. L., Ward, T., & Beaty, R. E. (2023). Divergent semantic integration (DSI): Extracting creativity from narratives with distributional semantic modeling. *Behavior Research Methods*, 55, 3726–3759. <https://doi.org/10.3758/s13428-022-01986-2>
- Kasof, J. (1997). Creativity and breadth of attention. *Creativity Research Journal*, 10(4), 303–315. https://doi.org/10.1207/s15326934crj1004_2
- Kaufman, J. C. (2019). Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 187–192. <https://doi.org/10.1037/aca0000217>
- Kempf, A., Benedek, M., & Schiavio, A. (2024a). An observation of a negative effect of social cohesion on creativity in musical improvisation. *Scientific Reports*, 14, 2922. <https://doi.org/10.1038/s41598-024-52350-7>
- Kempf, A., Maes, P.-J., Gener, C., & Schiavio, A. (2024b). Individual differences in music-induced interpersonal synchronization and self-other integration: The role of creativity and empathy. *Royal Society Open Science*, 11(11), 240654. <https://doi.org/10.1098/rsos.240654>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge University Press.
- Lebuda, I., & Benedek, M. (2023). A systematic framework of creative metacognition. *Physics of Life Reviews*, 46, 161–181. <https://doi.org/10.1016/j.plrev.2023.07.002>
- Lenth, R. (2025). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.11.0). <https://doi.org/10.32614/CRAN.package.emmeans>

- Lu, K., Gao, Z., Wang, X., Qiao, X., He, Y., Zhang, Y., & Hao, N. (2023). The hyper-brain neural couplings distinguishing high-creative group dynamics: An fNIRS hyperscanning study. *Cerebral Cortex*, 33(5), 1630–1642. <https://doi.org/10.1093/cercor/bhac161>
- Lucznik, K., May, J., & Redding, E. (2021). A qualitative investigation of flow experience in group creativity. *Research in Dance Education*, 22(2), 190–209. <https://doi.org/10.1080/14647893.2020.1746259>
- Lüdecke, D. (2024). *_sjPlot: Data Visualization for Statistics in Social Science_*. R package version 2.8.16, <https://CRAN.R-project.org/package=sjPlot>.
- Mednick, S. A. (1968). The remote associates test. *The Journal of Creative Behavior*, 2(3), 213–214. <https://doi.org/10.1002/j.2162-6057.1968.tb00104.x>
- Mendelsohn, G. A. (1976). Associative and attentional processes in creative performance. *Journal of Personality*, 44(2), 341–369. <https://doi.org/10.1111/j.1467-6494.1976.tb00127.x>
- Mockros, C. A., & Csikszentmihalyi, M. (1999). The social construction of creative lives. In A. Montuori & R. Purser (Eds.), *Social creativity* (Vol. 1, pp. 175–218). Hampton Press.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS One*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nakagawa, S., Johnson, P., & Schielzeth, H. (2017). The coefficient of determination R² and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14, <https://doi.org/10.1098/rsif.2017.0213>
- Niu, W., & Sternberg, R. J. (2006). The philosophical roots of western and eastern conceptions of creativity. *Journal of Theoretical and Philosophical Psychology*, 26, 18–38. <https://doi.org/10.1037/h0091265>
- Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in psychology. *Review of General Psychology*, 23(4), 403–424. <https://doi.org/10.1177/1089268019883821>
- Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. Ph.D. thesis, Department of Computing, City University, London, London, UK.
- Perry-Smith, J. E., & Challey, C. E. (2003). The social Side of creativity: A static and dynamic social network perspective. *The Academy of Management Review*, 28(1), 89–106. <https://doi.org/10.2307/30040691>
- Puente-Diaz, R. (2015). Creative self-efficacy: An exploration of its antecedents, consequences, and applied implications. *The Journal of Psychology*, 150(2), 175–195. <https://doi.org/10.1080/00223980.2015.1051498>
- Puente-Diaz, R., & Cavazos-Arroyo, J. (2018). Creative metacognitive feelings as a source of information for creative self-efficacy, creativity potential, intrapersonal idea selection, and task enjoyment. *Journal of Creative Behavior*, 54(3), 499–507. <https://doi.org/10.1002/jocb.384>
- Razumnikova, O. M. (2013). Divergent Versus Convergent Thinking. In E. G. Carayannis (Ed.), *Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship*. Springer. https://doi.org/10.1007/978-1-4614-3858-8_362
- R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2025). *psych: Procedures for psychological, psychometric, and personality research* (R package version 2.5.6). Northwestern University. <https://CRAN.R-project.org/package=psych>
- Reybrouck, M. (2006). Musical creativity between symbolic modeling and perceptual constraints: The role of adaptive behaviour and epistemic autonomy. In I. Deliège & G. A. Wiggins (Eds.), *Musical creativity: Multidisciplinary research in theory and practice* (pp. 16–35). Psychology Press.
- Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology*, 46(2), 168–175. <https://doi.org/10.1177/0098628319834381>
- Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology*, 13, 913219. <https://doi.org/10.3389/fpsyg.2022.913219>
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24(1), 66–75. <https://doi.org/10.1080/10400419.2012.652929>
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24, 92–96. <https://doi.org/10.1080/10400419.2012.650092>
- Runco, M. A., & Yorum, S. (2014). The neuroscience of divergent thinking. *Activitas Nervosa Superior*, 56(1), 1–16. <https://doi.org/10.1007/BF03379602>
- Saretzki, J., Forthmann, B., Benedek, M., Andrae, R., Meinzer, N., Bartlmä, L., & Braun, F. (2024). A systematic quantitative review of divergent thinking assessments. *Psychology of Aesthetics, Creativity and the Arts*.
- Sawyer, K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music*, 34, 148–165. <https://doi.org/10.1177/0305735606061850>
- Sawyer, K. (2012). Extending sociocultural theory to group creativity. *Vocations and Learning*, 5(1), 59–75. <https://doi.org/10.1007/s12186-011-9066-5>
- Sawyer, K., & De Zutter, S. (2009). Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 81–92. <https://doi.org/10.1037/a0013282>
- Schiavio, A., & Benedek, M. (2020). Dimensions of musical creativity. *Frontiers in Neuroscience*, 14, 578932. <https://doi.org/10.3389/fnins.2020.578932>
- Schiavio, A., Biasutti, M., Kempf, A., Popescu, T., & Benedek, M. (2024). The processes and relationships in composers scale. Construction and psychometric analysis of a new self-assessment inventory. *Music Perception*, 41(3), 217–231. <https://doi.org/10.1525/mp.2024.41.3.217>
- Schiavio, A., Gesbert, V., Reybrouck, M., Hauw, D., & Parncutt, R. (2019). Optimizing performative skills in social interaction. Insights from embodied cognition, music education, and sport psychology. *Frontiers in Psychology*, 10, 1542. <https://doi.org/10.3389/fpsyg.2019.01542>

- Schiavio, A., Moran, N., Antović, M., & van der Schyff, D. (2022c). Grounding creativity in music perception? A multidisciplinary conceptual analysis. *Music & Science*, 5. <https://doi.org/10.1177/20592043221122949>
- Schiavio, A., Moran, N., van der Schyff, D., Biasutti, M., & Parncutt, R. (2022b). Processes and experiences of creative cognition in seven western classical composers. *Musicae Scientiae*, 26(2), 303–325. <https://doi.org/10.1177/1029864920943931>
- Schiavio, A., & Nijs, L. (2022). Implementation of a remote instrumental music course focused on creativity, interaction, and bodily movement. Preliminary insights and thematic analysis. *Frontiers in Psychology*, 13, 899381. <https://doi.org/10.3389/fpsyg.2022.899381>
- Schiavio, A., Ryan, K., Moran, N., van der Schyff, D., & Gallagher, S. (2022a). By myself but not alone. Agency, creativity, and extended musical historicity. *Journal of the Royal Musical Association*, 147(2), 533–556. <https://doi.org/10.1017/rma.2022.22>
- Silvia, P. J., Wigert, B., Reiter-Palmon, R., & Kaufman, J. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, 6, 19–34. <https://doi.org/10.1037/a0024071>
- Simonton, D. K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford University Press.
- Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, 77, 184–197. <https://doi.org/10.1016/j.chb.2017.08.038>
- van der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, T. (2018). Musical creativity and the embodied mind. Exploring the possibilities of 4E cognition and dynamical systems theory. *Music & Science*, 1, <https://doi.org/10.1177/2059204318792319>
- Von Held, F. (2012). *Collective Creativity: exploring creativity in social network development as part of Organizational Learning*. Springer.
- Walton, A., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., & Richardson, M. J. (2017). Creating time: Social collaboration in music improvisation. *Topics in Cognitive Science*, 10, 95–119. <https://doi.org/10.1111/tops.12306>
- Wiggins, G. A. (2016). Defining inspiration? Modelling the non-conscious creative process. In D. Collins (Ed.), *The act of musical composition: Studies in the creative process* (pp. 233–255). Routledge.

Appendix

Table A1. Background questionnaire.

Question	Response
Age	[Free-text input]
Gender	Female, Male, Prefer not to say, [Free-text input]
What is your native language?	[Free-text input]
What is your favorite musical genre to listen to?	[Free-text input]
Have you ever written the lyrics of a song?	Yes, No
If yes (you have written the lyrics of a song), can you tell us more about this experience?	[Free-text input]
Please choose on word to describe your mood right now.	[Free-text input]
Do you consider yourself an empathic person?	Yes, No
Do you often find yourself mind-wandering?	Yes, No

Table A2. Post-questionnaires presented after each lyric writing trial. Questions with a (*) were only presented after the second condition.

Question	Response	In condition (OTHER)	In condition (SELF)
Please rate your enjoyment of the music track used in the experiment.	Likert – [1/I did not like it at all. 7/I fully enjoyed.]	x	x
Please estimate how many times you listened to the track before writing lyrics.	[Free-text input]	x	x
Did you listen to the track also while writing lyrics?	Yes, No	x	x
If yes (you listened to the track while writing lyrics) please estimate how many times you listened to the track while writing lyrics.	[Free-text input]	x	x
Did you sing along (e.g., humming) while listening to the track?	Yes, No	x	x
Did you take notes (e.g., sketching the lyrics) while listening to the track?	Yes, No	x	x
Please estimate how long it took you to complete writing the lyrics (in minutes).	[Free-text input]	x	x
How difficult was it to write the lyrics?	Likert – [1/Not difficult at all. 7/Extremely difficult.]	x	x
How well did you manage to think of someone else as the singer you wrote the lyrics for?	[1/Not at all well. 7/Perfectly well.]	x	
Who was the imagined person for whom you wrote the lyrics? Please describe.	[Free-text input]	x	
Which picture best describes your relationship with the imagined person (other)?	IOS Scale [1, 7]	x	
How creative would you rate your final set of lyrics?	Likert – [1/Not at all creative. 7/Exceptionally creative.]	x	x
How happy are you with your final set of lyrics?	Likert – [1/Not at all happy. 7/Fully satisfied.]	x	x
Please write down your lyrics without dividing syllables.	[Free-text input]	x	x
How well did you manage to think of yourself as the singer who you wrote the lyrics for?	Likert – [1/Not at all well. 7/Perfectly well.]		x
(*) Did you experience a difference between writing lyrics for yourself and writing lyrics for somebody else?	Yes, No	x	x
(*) If yes (you experienced a difference between these modalities), please tell us more about this experience.	[Free-text input]	x	x

Table A3. Questionnaire on which expert songwriters evaluated lyrics.

Question	Response
The lyrics go well with the music (i.e., can be sung efficiently).	Likert – [1 (Strongly disagree), 7 (Strongly agree)]
The lyrics are coherent with the music.	Likert – [1 (Strongly disagree), 7 (Strongly agree)]
Overall, I find the lyrics creative.	Likert – [1 (Strongly disagree), 7 (Strongly agree)]
The lyrics are original.	Likert – [1 (Strongly disagree), 7 (Strongly agree)]
The lyrics are surprising.	Likert – [1 (Strongly disagree), 7 (Strongly agree)]
Overall, I like the lyrics.	Likert – [1 (Strongly disagree), 7 (Strongly agree)]