Evaluation of Poetic Creativity: Predictors and the Role of Expertise - A Multilevel Approach

Soma Chaudhuri[1], Maura Dooley[2], Dan Johnson[3], Roger Beaty[4], and Joydeep Bhattacharya[1]

[1]Department of Psychology, Goldsmiths, University of London

[2]Department of English and Creative Writing, Goldsmiths, University of London

[3]Department of Cognitive & Behavioural Science, Washington and Lee University

[4]Department of Psychology, Pennsylvania State University

Author Note

Joydeep Bhattacharya: https://orcid.org/0000-0003-3443-9049

Soma Chaudhuri: https://orcid.org/0000-0003-0501-4519

Dan Johnson: https://orcid.org/0000-0001-8312-6191

Roger Beaty: https://orcid.org/0000-0001-6114-5973

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Soma Chaudhuri, Department of Psychology, Goldsmiths, University of London, New Cross, SE14 6NW. Email: schau002@gold.ac.uk

Evaluation of Poetic Creativity: Predictors and the Role of Expertise - A Multilevel Approach

Abstract

Poetry is one of the most creative expressions of language, but how we evaluate the creativity of a poem is not properly characterized. The present study investigated the role of various subjective qualities – clarity, aesthetic appeal, felt valence, arousal, and surprise – in predicting the creativity judgment of English poems. Participants (N=129) were presented with a broad range of English poems; they rated each poem on six characteristics: clarity, aesthetic appeal, felt valence, felt arousal, surprise and overall creativity. Linear multilevel analysis showed that aesthetic appeal was the strongest predictor of poetic creativity, followed by surprise and felt valence. Multilevel mediation analysis indicated significant mediation by surprise and felt valence on the relationship between aesthetic appeal and creativity at both within and between-participant levels. Further, expertise in English literature was found to significantly moderate the effects of all three predictors on the evaluation of creativity. The study simultaneously captured the surprise-evoking line(s). Using the semantic distance computing approach, we have shown the objective validation of the subjectively chosen line(s) of surprise. Altogether, our findings suggest a parsimonious model of evaluation of creativity of poems and its interaction with expertise.

Keywords: poetry, creativity, evaluation, expertise

Evaluation of Poetic Creativity: Predictors and the Role of Expertise -A Multilevel Approach

Poetry is one of the most creative expressions of language. It uniquely captivates readers, encouraging them to perceive unusual and complex issues. Poetry evokes strong emotion (Wassiliwizky et al., 2017) and its reception is associated with aesthetic pleasure. The psychology of creativity and aesthetics have been extensively studied using stimuli like paintings (e.g., Cupchik et al., 2009; Hagtvedt et al., 2008), music (e.g., Zioga et al., 2020; Jackendoff & Lerdahl, 2006; Koelsch, 2014), films (e.g., Hanich et al., 2014; Plucker et al., 2009). However, the factors used to judge a poem's creativity are largely unknown. Empirical research on the influence of the emotional impact of poetic language and the associated aesthetic pleasure in the reception of poetry has not been studied much to evaluate poetic creativity.

Most research on poetry has focused on its objective features like textual structure, i.e., rhythm, rhyme, meter (Obermeier et al., 2013; Lau et al., 2018), phonological constructs (Aryani et al., 2016), metaphors (Rasse et al., 2020) and subjective features such as expertise (Kaufman et al., 2008) and individual differences (Belfi et al., 2018; Hitsuwari & Nomura, 2022a , 2022b). Using *SciFaiku* (a form of haiku based on science fiction), Kaufman et al., (2010) compared overall creativity ratings by sex and ethnicity of both writers and raters. Their findings suggested that poems written by females were judged as more creative and female raters were more consistent compared to their male counterparts. A recent study on haiku and sonnets has shown that vividness of imagery is the best predictor of aesthetic appeal, followed by perceived valence and arousal (Belfi et al., 2018). In another study, felt

valence and imagery vividness in haiku predict its aesthetic appeal, where felt valence partially mediates the influence of imagery vividness on aesthetic appeal (Hitsuwari & Nomura, 2022b). Recent brain imaging studies on poetry appreciation have explored the emotional impact of poetic language and the associated aesthetic pleasure, which engage brain areas of primary reward (Wassiliwizky et al., 2017). A comprehensive neuroscientific study on poetry composition examined the assessment process, product, and expertise in a single experiment to characterize the neural mechanisms of both the generation and evaluation phases of creativity (Liu et al., 2015); the medial prefrontal cortex was activated during both phases, while the dorsolateral prefrontal and parietal cortex were activated in a phase-dependent manner, and experts showed higher deactivation in these regions. Altogether, these findings highlight a dynamical interplay between motivation, cognitve control related brain regions and their connectivities with multiple brain regions during creative behaviour.

The evaluation of the creativity of poems has not been studied much so far to explore the potential predictors influencing judgments of poetic creativity per se. In this study, we addressed how people would evaluate poems and their creativity, which subjective qualities would predict their overall creativity judgment, and how expertise would moderate the influence of the prospective predictors. We focused on five predictors: three factors associated with poetry - clarity, aesthetic appeal, and surprise, and two associated with the emotions felt after reading each poem – valence and arousal. The following sections briefly review the role of potential contributory factors in predicting the creativity of literary art forms, focusing on poetry and its evaluation.

Clarity, Aesthetic Appeal, and Creativity

Clarity of a text refers to a clear, understandable, comprehensible piece of writing which can effectively communicate with its readers. Clarity is perhaps the primary quality of a good style of any valued written communication, poetic or transactional. In poetry, the poet

must transfer his/her thoughts into words to effectively interact with readers. Considering that clarity is one of the fourteen dimensions of the judgment of the creativity of a poem (Amabile, 1982), we proposed that clarity would show a positive influence on the judgment of poetic creativity, i.e., the higher the comprehension of the poem, the better the assessment score for the creativity (*Hypothesis 1*).

Evaluating a piece of art must involve aesthetic appreciation, frequently termed as aesthetic appeal. Aesthetic evaluation of art has been studied in the context of visual art (Hagtvedt et al., 2008), music (Belfi, 2019; Brattico et al., 2017), and poetry (Kraxenberger & Menninghaus, 2017; Scharinger et al., 2022; Belfi et al., 2018; Hitsuwari & Nomura, 2022b, 2022c ; Jacobs, 2017). Aesthetic appreciation of poems was mostly studied so far to explore how it depends on the content of poems (Scharinger et al., 2022), personality traits, and psychological states such as valence, arousal vivid mental imagery (Belfi et al., 2018; Hitsuwari & Nomura, 2022b). However, how a poem's aesthetic appeal influences the poem's creativity judgment is not yet known. Here, we predicted that the aesthetic appeal of a poem would be a significant predictor of poetic creativity; poems with higher aesthetic appeal would receive higher creativity ratings (*Hypothesis 2a*). Furthermore, as the emotional content of verbal materials affects reading and a subjective emotional appraisal may play a critical role in aesthetic experience (Chatterjee & Vartanian, 2014; Leder et al., 2004), we expected both felt valence and arousal would partially mediate the relationship between aesthetic appeal and creativity of poem (*Hypothesis 2b* & *Hypothesis 2c* respectively). Finally, considering surprise as an interest-evoking construct associated with aesthetic appreciation (Silvia, 2009), we expected a partial mediation by surprise in explaining the process through which aesthetic appeal and creativity were related (*Hypothesis 2d*).

Emotions and Creativity

Poetry evokes emotions. The celebrated American poet Robert Frost once said, *"A complete poem is one where an emotion has found its thought, and the thought has found words"* (Frost, 1963). The semantic contents of a poem, along with its prosodic cues like meter, rhyme, rhythm, and aesthetics, evoke emotions (Johnson-Laird & Oatley, 2022). The emotional response to poetry is both a process initiated by the poet and a reciprocal process undertaken by the reader. The evaluation of ideas is an active interaction between the evaluator and the product dependent on the evaluator's emotional state (Mastria et al., 2019). Poetry has shown highly pleasurable emotional effects eliciting peak emotional experiences, including chills and goosebumps (Wassiliwizky et al., 2017). We expected both dimensions of felt emotion, i.e., valence (*Hypothesis 3a*) and arousal (*Hypothesis 3b*), to be significant positive predictors of poetic creativity. Of note, here we focused on the felt emotion, i.e., the emotion experienced by the participants induced by the poem.

Surprise and Creativity

Surprise has been found as a stronger predictor of creativity than value after controlling for originality (Acar et al., 2017). It supports the 3-criterion definition of creativity (Simonton, 2012). Poetry provides readers with unexpected shifts of concepts or violations of expectations, creating elements of surprise. Notably, surprise is a key mechanism by which music induces emotion (Juslin & Västfjäll, 2008) and predicts musical pleasure (Cheung et al., 2019). Moreover, the way to understand the effect of music is to focus on this 'kinetics' of expectation and surprise (Meyer, 1967). Therefore, we predicted that evaluation of creativity of a poem would depend on how readers would experience the expectancy violation in the context of the poem. We expected that surprise would be a significant positive predictor of poetic creativity; the more surprising the poem is, the more creative it is (*Hypothesis 4a*).

In a semantic space, the more unrelated the two concepts are, the more novel or creative the new concept will be (Kenett, 2019). In the present study, along with the surprise evaluation scores of the poems, we simultaneously captured the subjectively chosen line(s) of surprise. We computed the semantic relatedness of the surprise-evoking line (s) with respect to the two preceding lines. For this, we used SemDis, an automated scoring approach of verbal creativity that uses natural language processing to quantify the semantic relatedness of texts (Beaty & Johnson, 2021). We expected that semantic unrelatedness of the subjectively chosen line(s) would predict surprise scores of the poems. Specifically, the more unrelated the lines are, the more surprising the content will be. Therefore, we predicted that the subjectively chosen surprise-evoking line(s) would have some objective validation (*Hypothesis 4b*).

Role of Expertise in Poetry Evaluation

Art-trained and naïve participants perceptually explore artworks differently (Winston & Cupchik, 1992; Bhattacharya & Petsche, 2005). There are significant differences in aesthetic appreciation as a function of people's experience and knowledge of art (Cela-Conde et al., 2011). A neurophysiological study suggests that the non-experts show a higher emotional reaction in arousal, whereas the experts are more cognitively engaged with the same stimuli (Cartocci et al., 2021). Physiological correlates of art appreciation using facial EMG have shown that aesthetic expertise fosters a detached mode promoting an attenuation of the impact of emotional content (Leder et al., 2014). In a study on the creativity judgment of poetry, expert raters rated the poems as less creative than novice raters with differential levels of interrater agreement (Kaufman et al., 2008). In the present study, we expected significant moderation of expertise on aesthetic appeal, valence, arousal, and surprise in predicting poetic creativity. The likelihood of style-related processing is greater with higher expertise (Augustin & Leder, 2006). As aesthetics is a style-based perceptual construct of

poetry, we predicted that experts with formal knowledge of English literature would experience a stronger positive impact of aesthetic appeal on creativity ratings compared to the naïve participants (*Hypothesis 5a*). Further, in line with previous research, we expected that the positive impact of both felt valence and felt arousal on creativity scores would be less pronounced in experts compared to nonexperts (*Hypotheses 5b* and *5c* respectively). Surprise is an interruption mechanism and a short-lived mixed emotion (Meyer et al., 1997). It is evoked by unexpected events which might interrupt ongoing thoughts and motivate people to pay attention to the unexpected stimulus (Noordewier & Breugelmans, 2013). Therefore, our specific prediction was that the positive impact of surprise on creativity scores would be more pronounced in nonexperts (*Hypothesis 5d*). We assumed that nonexpert individuals may be more intrigued by unexpected elements of a poem. This, in turn, could result in an enhanced ability to appreciate and assess creativity when experienced surprising elements within poems.

Method

Participants

Using G*power (version 3.1; Faul et al., 2007), we found that a minimum sample size of 92 was required to achieve an 80% power for a medium effect ($f^2 = 0.15$) with squared multiple correlation $\rho^2 = 0.13$. We recruited 129 healthy adult participants via Prolific®. Since the task required approximately one hour to complete, we excluded 30 participants who exceeded the time limit of two hours. Three participants were excluded from the analyses due to their identical responses on the subjective rating measures across the poems. Our final sample size (N = 96) was, therefore, adequate in terms of statistical power. This sample consisted of 32 males, 63 females, and one preferred not to say; the average age of the participants was 31.94 years with a standard deviation of 13.09.

Considering that some formal training and experience in the target domain is necessary for judges (Amabile, 2018; see also Kaufman et al., 2009), we considered participants with a formal academic degree in English literature as the objective criterion for expert selection. The rationale behind our expert-selection criterion was also supported by research in other art domains like music and visual art, where experts were chosen with formal degrees in relevant disciplines (Bhattacharya & Petsche, 2005; Kottlow et al., 2011; Fudali-Czyż et al., 2018). Therefore, participants holding a formal degree (Bachelor's degree or above) in English literature were considered as experts. Thirty-nine participants were assigned as experts. The remaining 57 participants were considered as nonexperts. All participants provided informed consent before the data collection and were paid £7.50 per hour as monetary incentive.

Stimuli

Thirty-six original English language poems from wide-ranging structures and contents (see Table 1 for details of the poems), lines and words (mean number of lines = 11, SD = 3.24; mean word count = 71.25, SD = 28.99) were selected as stimuli. Previous studies on the aesthetic evaluation of poems were based primarily on haiku, sonnet (Belfi et al., 2018; Hitsuwari & Nomura, 2022b), or Shakespeare's sonnet (Papp-Zipernovszky et al., 2021). Despite offering some advantages, like brevity and structural rigidity/consistency, these poems may not be appropriate representatives of varieties of English poems. Therefore, we did not restrict our choice of stimuli to a particular genre or form.

We initially selected 108 poems from various popular online poetry resource sites, such as Poetry.org (http://www.poetry.org/), Poetry Foundation (https://www.poetryfoundation.org/), and Academy of American Poets (https://poets.org/). One of the co-authors (M.D.), an award-winning poet and a senior professor in creative writing and English, rated these poems on surprise and creativity on a 7-point Likert scale

from 1 (“Extremely Low”) to 7 (“Extremely High”). We finally selected 36 poems based on their surprise scores: 18 low-surprise poems (with ratings of 4 or lower) and 18 high-surprise poems (with ratings of 6 or above).

----------[Insert Table 1 here]----------

Procedure

The survey link was distributed via Prolific®, and all data was collected online via Qualtrics®. First, participants received a broad overview of the study, followed by instructions about rating the poems. A sample poem was provided at the outset to make the evaluation process easier to understand. Next, each poem was presented for 30 seconds. The titles of the poems and the respective poets were intentionally withheld to prevent any potential bias towards specific poems or poets. Subsequently, the participants rated the poems on six dimensions in the following order: clarity, aesthetic appeal, felt valence, felt arousal, surprise, and creativity. Each rating was made on a 7-point Likert scale from 1 (“Extremely Low”) to 7 (“Extremely High”).

Additionally, for each poem, participants were asked to identify the specific line(s) within each poem that they found most surprising or unexpected. Finally, after rating 36 poems, participants were requested to provide their demographic details, including gender, age, ethnicity, highest educational qualifications, association with English poetry, and affinity towards reading and writing English poetry. It took an average of one hour to complete the task. The study protocol was approved by the local Ethics Committee of the Department of Psychology.

Analysis

The primary aim of our study was to investigate which of the five subjective ratings - clarity, aesthetic appeal, felt valence, arousal, and surprise - would best predict the overall creativity judgment of poems. General data visualization and checks comprised descriptive

statistics of the variables, normality checks for the outcome variable, multicollinearity checks for the independent variables, and the internal consistency check. Results confirmed normally distributed outcome variable with no multicollinearity within independent variables (VIF<3). Cronbach's alpha (0.87) and McDonald's omega (Omega hierarchical = 0.77) confirmed internal consistency across items.

Our experimental data consisting of 3456 responses (96 participants x 36 poems x 6 ratings) had the common multilevel structure; responses (Level-1) being nested within participants (Level-2). Hence, we considered a linear mixed effects model to explore the variability in the between-subject evaluations and within-subject relationships. The null model demonstrated a substantial 54% variance explained by the grouping variable (participants), which supported the use of a linear mixed model over standard regression models to account for the multilevel structure of the data. Further, the intraclass correlation coefficient (ICC = 0.28) indicated that the Level-1 dependent variable (creativity) was not independent of the Level-2 grouping variable (participants). Therefore, linear mixed modeling was proven justified. To identify the most effective predictors in determining the overall creativity judgment of poems, five separate maximum likelihood linear mixed models using the *lme4* package (Bates et al. 2015) in R (version 4.0.3) were run on creativity scores. Five potential predictors were centred within each subject (i.e., group mean-centred) before entering the model to obtain an unambiguous estimate of the within-group effect (Enders & Tofighi, 2007). Five potential predictors and their interactions with expertise were considered as fixed effects, and the intercepts for participants as random effects. The best model fit results identified the potential predictors of poetic creativity and their interactions with expertise. Furthermore, we investigated the mediation effect of variables on the relationship between the best predictor and creativity. For this, we performed a multilevel

mediation analysis with all variables measured in Level-1, using MLMED, a computational Macro for SPSS (Rockwood, 2017; Hayes & Rockwood, 2020).

Finally, we explored objective validation of the subjectively chosen line(s) of surprise. We used SemDis (Beaty & Johnson, 2021) to quantify the semantic unrelatedness of the chosen lines with the preceding context. Two lines preceding the surprise-evoking line(s) chosen by the participants were considered as the reference. We computed the mean SemDis scores using a multiplicative compositional model. This model creates a single vector for a phrase by taking the product of all word vectors allowing the shared semantic dimensions of the component words to get higher scores in the final vector and the unshared semantic dimensions between words to get lower scores (Beaty & Johnson, 2021). We have chosen the multiplicative model as it outperforms the additive model in its correlation with human ratings of relatedness and creativity (Beaty & Johnson, 2021; Mitchell & Lapata, 2010). It also substantially mitigates the elaboration bias demonstrated in prior research using semantic distance to capture creativity (Forthmann et al., 2019).

Results

Descriptive Statistics

Descriptive statistics and bivariate correlations of prospective predictors are shown in Table 2 and Table 3, respectively. The variables exhibit a slightly left-skewed distribution, as indicated by their negative skewness values approaching zero. Additionally, the near-zero kurtosis values of the variables confirm their approximate normal distribution. Variation Inflation Factors (VIF<3) confirm that there is no significant issue of multicollinearity among the variables (see Table 2). Bivariate correlations depict that creativity is positively and significantly correlated (all $p<.01$) with all five predictor variables: clarity (r =.42), aesthetic appeal (r = .66), felt valence (r =.60), arousal (r =.47), surprise (r =.52) (see Table 3).

----------[Insert Table 2 and Table 3 here]----------

Selection of Predictors

We used the forward selection method to include variables in the predictor model. The variable with the highest correlation with the outcome variable (creativity) was entered first into the null model, followed by other variables in the descending order of their correlations with creativity. So, the predictor variables were entered into the model in the following order: aesthetic appeal, felt valence, surprise, arousal, and clarity. The Information Criteria (e.g., AIC, BIC), the proportion of variance explained by fixed effects (R^2) and Likelihood ratio test statistic ($\Delta\chi^2$) were used to compare five linear mixed models. The model comparison results showed that the model comprising aesthetic appeal, felt valence, and surprise was the most parsimonious model-fit ($\Delta\chi^2$= 289.5, *BIC*= 9081.5, R^2= 0.34, p<0.001). Model comparison and selection of the best model-fit are shown in Table 4.

Aesthetic appeal was found to be the best predictor (b = 0.31, *SE* = 0.02, t =15.16, p<0.001) followed by surprise (b = 0.25, *SE* = 0.02, t = 14.72, p<0.001) and felt valence (b = 0.20, *SE*=0.02, t =10.27, p<0.001) supporting our Hypotheses 2a, 4a and 3a respectively. We observed a substantial increase in the ICC for the best-fit model (0.43) from that of the null model (0.28). Clarity was found to be a nonsignificant predictor (b=0.02, *SE* = 0.02, t = 0.97, p = 0.33) which did not support our Hypothesis 1 and hence was discarded. Furthermore, while arousal exhibited significance (b = 0.10, *SE* = 0.02, t = 4.75, p < 0.001), its inclusion did not enhance parsimony of the model (see Model 4 in Table 4), leading to its exclusion as a viable predictor. This outcome did not support our Hypothesis 3b. The linear mixed model result for the best model-fit is shown in Table 5.

However, before discarding clarity and arousal as potential predictors, we performed partial correlation analyses with them as confounding factors. The partial correlation between aesthetic appeal and creativity, controlling for clarity (r = 0.42) and controlling for arousal (r

= 0.37), suggested a negligible effect of clarity and arousal on the strength of the relationship between aesthetic appeal and creativity.

----------[Insert Table 4 and Table 5 here]----------

Mediation by Felt Valence and Surprise

We examined felt valence and surprise as the mediators in the same model using parallel (multilevel) mediation. We assumed that no mediator causally influences each other. Considering that felt valence and surprise may have different implications, we independently examined their pathways to creativity, hypothesizing that both felt valence and surprise would mediate the relationship between aesthetic appeal and creativity (*Hypotheses 2b* and *2d,* respectively). As arousal has not been considered a potential predictor, we did not deploy arousal as a mediator.

The within-level indirect effects of aesthetic appeal on creativity via felt valence (b = 0.12, SE = 0.01, z = 11.27, 95% CI [0.1, 0.14]) and surprise (b = 0.1, SE = 0.1, z = 14.62, 95%CI [0.08, 0.11]) were statistically significant with a proportion of mediation 26.15% and 21.91% respectively. The between-level indirect effects of aesthetic appeal on creativity via felt valence (b = 0.01, SE = 0.01, z = 0.08, 95% CI [-0.17, 0.18]) and surprise (b = 0.12, SE = 0.05, z = 2.45, 95% CI [0.03, 0.23]) with a proportion of mediation 0.88% and 13.29% respectively. Hence, the relationship between aesthetic appeal and creativity was partially mediated by felt valence within levels and by surprise across levels (see Figure 1). The multilevel mediation analysis was performed using the MLMED macro in SPSS (Rockwood, 2017; Hayes & Rockwood, 2020).

----------[Insert Figure 1 here]----------

Moderating Role of Expertise

Considering a formal degree in English literature as the objective criterion of expertise, we found a statistically significant expertise-moderated effect on all three potential

predictors of poetic creativity (see Figure 2). Simple slopes analyses result (see Table 6) show that in the case of the relationship between aesthetic appeal and creativity, slope for nonexperts is: b = 0.31, SE = 0.02, 95% CI = [0.27, 0.35], and for experts it is: b = 0.38, SE = 0.02, 95% CI = [0.34, 0.43]. Non-overlapping intervals with a difference in slopes between nonexperts and experts (b = -0.07, z ratio = -2.33, p = 0.02) show a significant expertise moderated effect. For the relationship between felt valence and creativity, slope for nonexperts is: b = 0.20, SE = 0.02, 95% CI = [0.16, 0.24], and for experts it is: b = 0.12, SE = 0.02, 95% CI = [0.08, 0.16]. As the intervals do not overlap, there is a clear difference in slopes between nonexperts and experts (b = 0.08, z ratio = 2.80, p = 0.01), consequently showing a statistically significant expertise-moderated effect. In the case of the relationship between surprise and creativity, slope for nonexperts is: b = 0.25, SE = 0.02, 95% CI = [0.22, 0.29], and for experts it is: b = 0.19, SE = 0.02, 95% CI = [0.15, 0.23]. As there is almost no overlapping in the intervals, there is a difference in slopes between nonexperts and experts (b = 0.06, z ratio = 2.40, p = 0.02), consequently showing a statistically significant expertise moderated effect.

---------[Insert Table 6 and Figure 2 here]---------

Further, considering longer association with poetry as a potential subjective criterion for expertise, we examined how it would influence the assessment of poetic creativity. Twenty-one participants who reported being associated with English poetry for 10 years or more were included as experts. The results of slope analyses (see Supplementary: Table S1) showed clear overlapping confidence intervals and no statistically significant difference in slopes for the relationships between aesthetic appeal, felt valence, and surprise with creativity. Supplementary Table S2 illustrates the comparison of the expertise-moderated effects considering two expertise criteria for judgment of poetic creativity.

In our analysis, we excluded participants who took more than 2 hours to complete the task, yet it can be argued that deep contemplation of poems may require individualized self-paced engagement based on one's perceptive level. Therefore, to accommodate this possibility, which is also ecologically valid, we replicated the analyses with a larger sample of 126 participants while ignoring the time taken to complete the task using both criteria of expertise. With our primary objective criterion of expertise (i.e., participants (N=49) with formal degree in English literature as experts), the best model-fit result is as follows: aesthetic appeal (b = 0.32, SE = 0.02, t =18.81, p<0.001), felt valence (b = 0.20, SE = 0.02, t =12.40, p<0.001), and surprise (b = 0.24, SE= 0.01, t =16.44, p<0.001) significantly predicted poetic creativity. The best model-fit result with long association with poetry as expertise condition (N=32) is as follows: aesthetic appeal (b = 0.33, SE = 0.01, t = 22.97, p<0.001), felt valence (b = 0.17, SE = 0.01, t =12.74, p<0.001), and surprise (b = 0.23, SE = 0.01, t = 17.68, p<0.001) were the significant predictors of poetic creativity. Using our primary objective criterion of expertise, we found statistically significant expertise-moderated effects on all predictors in the creativity judgment of poems (see Table S3 in the Supplementary).

Objective Validation of Subjective Ratings of Surprise

Using a linear mixed model, we found a significant relationship between surprise scores (outcome variable) and SemDis scores (independent variable) with participants as the grouping variable (b = 1.22, SE = 0.07, t =17.43, p<0.001). This result supported Hypothesis 4b, indicating that the subjectively chosen line(s) of surprise had significant objective validation. Further, we conducted a sensitivity analysis using an alternative scoring method, Open Creativity Scoring (OCS: Organisciak & Dumas, 2020), with the semantic model approach. The results ensured the significant prediction of surprise ratings by OCS scores, (b = 2.18, SE = 0.10, t = 22.28, p <.001) (see Table 7).

----------[Insert Table 7 here]----------

Discussion

In the present study, we investigated the evaluation of the creativity of poems. Specifically, we explored the role of various subjective qualities – aesthetic appeal, clarity, felt valence, felt arousal, and surprise – in predicting the overall creativity judgment for a broad range of English poems. We observed that aesthetic appeal was the best predictor, followed by surprise and felt valence. Interestingly, clarity and arousal did not significantly contribute towards predicting creativity over and above the three earlier predictors. Furthermore, multilevel mediation analysis showed that felt valence and surprise significantly partially mediated the effect of aesthetic appeal on creativity within and between - participant levels. Expertise in English literature showed significant moderation in the relationship between creativity and all three predictors. Additionally, semantic unrelatedness between the surprise-evoking line (s) and the two preceding lines significantly predicted the surprise scores of the poems, revealing objective validation of the subjectively chosen line(s). We briefly discuss these findings and some limitations in the following spaces.

Aesthetic Appeal, Surprise, and Felt Valence Predict Poetic Creativity

Our most parsimonious model for predicting the creativity judgment of the poems consists of three predictors (in decreasing order of importance): aesthetic appeal, surprise, and felt valence. Aesthetic appeal has been studied earlier in poetry, exploring how psychological states like valence, vividness in imagery, and arousal predicted aesthetic appeal (Belfi et al., 2018; Hitsuwari & Nomura, 2022b). We explored how aesthetic appeal would predict the creativity of poems. We propose that an individual's subjective and intuitive perception of the creative potential of a poem is majorly judged by the appreciation of the aesthetic appeal of the poem. After aesthetic appeal, the second most important predictor was surprise. Surprise in a poem can be operationalized as a violation of expectancy in the

concept of the poem., While reading, a violation of expectancy in its concept and/or context might break the monotone of the poem and make the readers more cognitively engaged with interest and curiosity. In this regard, surprise might have acted as a novelty-detecting and interest-evoking construct in a poem while probing into poetic creativity. The third predictor in the most parsimonious model of the creative evaluation was felt valence. Poetry evokes emotion, and literature suggests that valence influences aesthetic appreciation (Leder & Nadal, 2014). Our study suggests that feeling of positive or negative valence positively influences readers' judgment, cognition, and receptivity towards novel perspectives of the poem, which further augment the creativity evaluation process.

Contrary to our hypothesis, although clarity exhibited a moderately high correlation with creativity (r =.42), it did not emerge as a significant predictor of creativity. Additionally, the partial correlation between aesthetic appeal and creativity, while controlling for clarity, supported the redundancy of clarity as a contributory parameter in judging the creativity of a poem. Clarity, distinct from readability, decreases with abstraction, technical language, and passive writing, whereas knowledge increases the use of these parameters (Warren et al., 2021). So, it might be argued that the creative evaluation of poetry, as a high-level cognitive construct, necessitates the indispensability of abstractness, technicality in language, and passive writing for its evaluation. Nevertheless, the way clarity was included in our study revealed it to be a relatively low-level construct for judging creativity of poetry. However, it remains unclear whether a sense of understanding or comprehension is essential for such judgment.

Further, despite having a reasonably high correlation with creativity (r = .47) and showing statistically significant results in the model building, arousal did not improve the model noticeably. In short, arousal did not improve the explanatory predictive power substantially. Our model selection aimed to minimize the number of predictors while

accounting for maximum variance in the criterion. The higher unexplained variance by arousal indicated a weaker strength of association with creativity. Furthermore, felt valence and arousal had shown a high correlation (r =.64), raising concerns about potential multicollinearity and its potential impact on the model's explanatory power.

We found aesthetic appeal to be the best predictor of the evaluation of poetic creativity. Aesthetic appeal of poems is influenced by various factors, including content (Scharinger et al., 2022), vivid mental imagery, personality traits, and psychological states of readers (Belfi et al., 2018; Hitsuwari & Nomura, 2022b). This study demonstrates for the first time the robust predictive power of aesthetic appeal in judging the creativity of poems.

Surprise was the second-best predictor of poetic creativity, supporting the 3-criterion definition of creativity with surprise as a key ingredient of creativity (Simonton, 2012). The abrupt change in context and/or theme of the poem might have cognitively engaged the readers and sparked their creative thinking towards the judgment of the creativity of poem. Hence, we propose that similar to other creative products, the evaluation of poetic creativity could be predicted by its surprise content. Additionally, the SemDis scores, which represent semantic unrelatedness, significantly predicted the subjective surprise ratings, therefore providing some objective validation of subjectively chosen surprise-evoking line(s). The results of the analyses using Open Creativity Scoring (OCS) further strengthens the validity of the results previously observed with SemDis scores. Therefore, it is justifiable to mention that the subjectively chosen line(s) were not selected randomly, and the surprise ratings can be objectively verifiable based on semantic unrelatedness.

Felt valence was a significant predictor of poetic creativity, supporting our hypothesis. We consider the judgment of a poem's creativity as a higher-order interpretation that occurs during later stages of information processing and often relies on the poem's representational content (Leder et al., 2012). Our result has shown that a positive emotional

valence would lead to a greater appreciation of poems, particularly in terms of poetry-specific ideas**,** concepts, and potentially the abstractness in the poem. This positive emotional valence further contributes to the evaluation of poetic creativity.

Felt Valence and Surprise Partially Mediate Aesthetic appeal-Creativity Relationship

The multilevel mediation result showed that the relationship between aesthetic appeal and the creativity of poems was partially mediated by felt valence and surprise, particularly at the within-participants level. The significant partial mediation by valence supports earlier studies, highlighting the critical role of subjective emotional appraisal in the aesthetic experience (Chatterjee & Vartanian, 2014; Leder et al., 2004). The significant partial mediation by surprise indicated that surprise, as an interest-evoking construct operationalized through the expectancy violation of readers, influenced the relationship between aesthetic appeal and creativity. Both mediators showed higher partial mediation effects at the within-participant level, demonstrating the variations in felt valence and surprise evaluations within the population. However, felt valence showed a higher partial mediation effect than surprise at the within-participants level, suggesting that aesthetic appeal predicted poetic creativity through the influence of valence. In other words, as the aesthetic appeal increased, individuals felt more positive valence while evaluating poem's creativity.

Expertise Moderates Creativity Judgment of Poetry

We found a significant moderating effect of expertise on the creativity ratings of poems. We selected experts based on an objective criterion of education in the field of English literature, i.e., experts possessed formal knowledge in the relevant domain. All three predictors of poetic creativity were found to be moderated by expertise, providing support for our hypotheses.

Simple slopes analyses (see Table. 6) showed that the positive effect of aesthetic appeal on creativity was stronger for experts. This indicates that aesthetic appeal had a more

pronounced positive effect on the judgment of creativity for experts with greater abilities in processing style-based information, including aesthetics. Of note, the effect size of the interaction was small, which could be attributed to two possible factors. Firstly, a relatively narrow range of expertise was tested in this study. If participants with a higher level of expertise in poetry, such as eminent poets, poetry critics, or academics in English literature, had been included, it might have led to stronger effects. Secondly, as experts are more likely to hold an art bias than nonexperts (Glăveanu, 2014), the contemplative appreciation of aesthetic appeal of the chosen poems might not have shown significant differences between experts and nonexperts. Therefore, these findings suggest a trend of overlap between domain-specificity and domain-generality in the aesthetic appreciation of poetry when evaluating its creativity.

We found a significant moderation effect of expertise on felt valence, supporting our hypothesis. Simple slopes analyses showed that the positive effect of felt valence on creativity was increasingly dampened for experts. Figure 2 shows a steeper gradient of the impact of felt valence on creativity ratings for nonexperts than experts. In other words, felt valence showed a more pronounced effect on creativity for nonexperts than experts. This finding aligns with earlier research showing an attenuated trend in valence ratings by experts compared to laypeople (Leder et al., 2014) and further supports the notion that poetry engages the minds of experts more intellectually while evoking stronger emotional responses in nonexperts (Cartocci et al., 2021). Considering that emotion is a robust predictor of liking for various art forms (Leder et al., 2012), it might be argued that nonexperts' judgment of emotional valence might involve their liking or preference. On the other hand, experts, driven by their knowledge-based skills, judged the creativity of poems from a more intellectual and cognitively comprehensive perspective. So, we observed a clear interplay between cognitive and emotional processes in experts and nonexperts while evaluating poetic creativity.

Although emotional experience plays a central role in aesthetic viewing (Chatterjee, 2003), our study suggests that it negatively influences the creativity judgment of experts for the high-level literary domain, like poetry, compared to nonexperts. It is important to note that in our study, participants rated their felt valence, not the perceived valence. If perceived emotion had been measured, experts could have considered it more a poetic construct than a self-representing construct for judging poetic creativity.

We observed a significant moderation effect of expertise on surprise, supporting our hypothesis. Simple slopes analyses show that the positive effect of surprise on creativity was increasingly dampened for the experts compared to nonexperts. This result supports our hypothesis that surprise, as an interruption mechanism and a short-lived emotion (Meyer et al., 1997) may disrupt the flow of thoughts in experts while evaluating poetry. In the domain of music, it has been shown that strictly deterministic and ordered patterns are perceived as boring, while random patterns are perceived as unstructured and featureless (Abdallah & Plumbley, 2009). Therefore, we speculate that an excessive increase of entropy and surprisingness in a poem might appear disorganised and unengaging to experts, leading to a reduced impact of surprise on their creativity judgment. In contrast, surprise had a more favourable impact on creativity judgment for nonexperts than experts. This result may seem somewhat incongruent with the 3-criterion definition of creativity (Simonton, 2012). While surprise emerged as the second-most significant predictor of poetic creativity, indicating a positive linear relationship that aligns with the 3-criterion definition of creativity, the expertise-moderated effects on surprise showed that this positive relationship was considerably stronger for nonexperts than experts. Nonexperts perceived surprise as a more impactful and influential parameter for evaluating poetic creativity than experts. Therefore, these findings would not undermine the consistency of the 3-criterion definition of creativity

but suggest that our experimental design exhibited greater consistency with this definition among nonexperts compared to experts.

Our study suggests that the perception of poetry varies among readers with varying levels of expertise. Experts tended to prioritize the processing of aesthetic appeal, focusing on skill-dependent artistic features of poetry. On the other hand, nonexperts placed greater emphasis on the emotional valence and surprise experienced while judging the creativity of the poem.

Considering poetry and music are known to evoke emotions and create aesthetic appeal for readers and listeners (Wassiliwizky et al., 2017; Juslin & Västfjäll, 2008), it is plausible to speculate that readers' experiences with poetry may share some similarities with the experiences of listeners. Studies suggest that even basic listening can lead to the development of musical knowledge and the emergence of 'experienced' listeners (Bigand & Poulin-Charronnat, 2006); Koelsch, 2014). Drawing on this idea, while interpreting the modest interaction of expertise, we can argue that nonexperts in our study might have had a basic habit of reading, which could have enabled them to evaluate the aesthetic appeal of poetry at par with the experts while judging poetic creativity.

In line with previous research, we considered a formal degree in English literature as our primary objective criterion for expert selection. However, we explored extended familiarity with poetry as a potential subjective criterion for expertise, where no expertise-moderated effect was observed. It is important to note that the generalizability of these findings may be limited, as there could be a potential influence of age, because longer association with poetry would be more common among older individuals. Additionally, the imbalance in group sizes, with a much smaller number of experts (N = 21) compared to nonexperts (N = 75), might have limited the statistical power to detect differences in slopes,

i.e., the moderation effect. Moreover, a broader sample of 126 participants, accounting for greater ecological validity, substantiated our predictor model of poetic creativity and the expertise criterion, confirming the strength of our findings.

Limitations and Future Scope of Work

The present study, however, is subject to some potential limitations. First, we measured the felt emotions of participants, i.e., the emotions the participants felt while reading a poem. This contrasts with the previous studies which measured perceived emotions - the emotions evoked by the stimuli, i.e., the emotional content of a poem (Aryani et al., 2016; Belfi et al., 2018). Perceived and felt emotions may be distinct (Gabrielsson, 2001; Marin & Bhattacharya, 2010) because, while reading, emotions evoked by the content of a poem may not always induce a feeling of identical emotions within readers. Future work could explore whether perceived emotion would predict the creativity of a poem. Moreover, since experts were more inclined towards the content and style of a poem rather than driven by felt emotions, it would be interesting to study whether experts would consider poetry-elicited emotion or perceived emotion as the determinant factor for judging poetic creativity. Second, our model did not include any trait-level component or personality features. Research suggests a kind of "taste typicality" of the aesthetic experience of ordinary scenes and objects (Chen et al., 2022); on the other hand, substantial individual differences exist in the aesthetical evaluation of poems. For example, for haiku and sonnets, individual differences in visual imagery abilities were found to moderately predict their aesthetic appeal (Belfi et al., 2018), and visual imagery ability, awe-proneness, and nostalgia-proneness were shown to predict haiku's aesthetic appeal (Hitsuwari & Nomura, 2022b). Hence, future work could study how participants' personality traits would influence the evaluation of poetic creativity. Third, our model focused on the influence of context or appeal-based characteristics of poetry without specifically exploring structure, rhythm, form, genre, or

other aspects that set poetry apart from other forms of writing. We did not restrict our experimental stimuli regarding length, rhythmic pattern, or specific forms, such as the sonnet, haiku, or limerick. The number of poems in our study was insufficient to explore the genre and form-specific effects. However, it is important to acknowledge that such poem-based aspects may affect the evaluation of poetic creativity. Fourth, we did not randomize the order of ratings. As each participant had to provide 6 ratings for each of the 36 poems, i.e., 216 responses in total, we wanted to keep the participants in a flow without perplexing them with the randomness of the order of questions. However, the potential order effect could not be ruled out either. Finally, we did not address "familiarity" in this study. Familiarity could be a potential bias in judgment as it enhances processing fluency, leading to preferences (Reber et al., 1998) and influencing perceptual characteristics (Goldinger et al., 1999). However, greater processing fluency also contributes to a better understanding of an art work's meaning (Lindell & Mueller, 2011). Future studies could consider investigating the moderating role of familiarity in predicting creativity judgment of poetry.

Our study explored the potential predictors of the creative evaluation of poetry. We found aesthetic appeal was the best predictor, followed by surprise and felt valence. Clarity and felt arousal were not included in our parsimonious model. The influence of aesthetic appeal on creativity was partially mediated via surprise and valence at both within - and between-participant levels. Additionally, we found a significant moderation effect of expertise on all three predictors, indicating that experts and nonexperts differed in how these predictors influenced the overall creativity ratings. Moreover, we showed that the subjectively chosen surprising line(s) of poems would be objectively validated based on semantic relatedness. However, it is important to note that our study did not include other poem-based constructs like structure, form, and genre, which may also impact the evaluation of poetic creativity. Overall, this study contributes to the field of creativity research,

particularly in exploring domain specificity and evaluation, which have been relatively understudied compared to domain-general creative generation.

Data Availability

Data and codes are available in Open Science Framework (OSF):

https://osf.io/rqxm5/?view_only=b5ce75bf5d924d0ba2373ccb3c647803

References

Abdallah, S., & Plumbley, M. (2009). Information dynamics: patterns of expectation and surprise in the perception of music. *Connection Science*, *21*(2–3), 89–117. https://doi.org/10.1080/09540090902733756

Acar, S., Burnett, C., & Cabra, J. F. (2017). Ingredients of creativity: Originality and more. *Creativity Research Journal*, *29*(2), 133–144. https://doi.org/10.1080/10400419.2017.1302776

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43,* 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

Amabile, T. M. (2018). *Creativity in context: Update to the social psychology of creativity*. Routledge.

Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts, 10,* 191–204. https://doi.org/10.1037/aca0000033

Augustin, D., & Leder, H. (2006). Art expertise: A study of concepts and conceptual spaces. *Psychology Science, 48*(2), 135.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48.

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods, 53*(2), 757–780. https://doi.org/10.3758/s13428-020-01453-w

Belfi, A. M. (2019). Emotional valence and vividness of imagery predict aesthetic appeal in music. . *Psychomusicology: Music, Mind, and Brain* , *29*((2-3)), 128–135.

https://doi.org/10.1037/pmu0000232

Belfi, A. M., Vessel, E. A., & Starr, G. G. (2018). Individual ratings of vividness predict aesthetic appeal in poetry. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(3), 341–350. https://doi.org/doi:10.1037/aca0000153

Bhattacharya, J., & Petsche, H. (2005). Drawing on mind's canvas: Differences in cortical integration patterns between artists and non-artists. *Human Brain Mapping*, *26*(1), 1–14. https://doi.org/https://doi.org/10.1002/hbm.20104

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100–130. https://doi.org/https://doi.org/10.1016/j.cognition.2005.11.007

Brattico, P., Brattico, E., & Vuust, P. (2017). Global sensory qualities and aesthetic experience in music. *Frontiers in Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fnins.2017.00159

Cartocci, G., Rossi, D., Modica, E., Maglione, A. G., Martinez Levy, A. C., Cherubino, P., Canettieri, P., Combi, M., Rea, R., Gatti, L., & Babiloni, F. (2021). Neurodante: Poetry mentally engages more experts but moves more non-experts, and for both the cerebral approach tendency goes hand in hand with the cerebral effort. *Brain Sciences*, *11*(3), 1–25. https://doi.org/10.3390/brainsci11030281

Cela-Conde, C. J., Agnati, L., Huston, J. P., Mora, F., & Nadal, M. (2011). The neural foundations of aesthetic appreciation. *Progress in Neurobiology*, *94*(1), 39–48. https://doi.org/https://doi.org/10.1016/j.pneurobio.2011.03.003

Chatterjee, A. (2003). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts*, *4*, 55–60. https://doi.org/10.1037/e514602010-003

Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences*, *18*(7), 370–375. https://doi.org/https://doi.org/10.1016/j.tics.2014.03.003

Chen, Y.-C., Chang, A., Rosenberg, M. D., Feng, D., Scholl, B. J., & Trainor, L. J. (2022). "Taste typicality" is a foundational and multi-modal dimension of ordinary aesthetic experience. *Current Biology*, *32*(8), 1837-1842.e3. https://doi.org/https://doi.org/10.1016/j.cub.2022.02.039

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology*, *29*(23), 4084-4092.e4. https://doi.org/https://doi.org/10.1016/j.cub.2019.09.067

Cupchik, G. C., Vartanian, O., Crawley, A., & Mikulis, D. J. (2009). Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition*, *70*(1), 84–91. https://doi.org/https://doi.org/10.1016/j.bandc.2009.01.003

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2019). Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior*, *53*(4), 559–575.

Frost, R. (1963). *The letters of Robert Frost to Louis Untermeyer* (L. Untermeyer (ed.)). Holt Reinhart and Winston. New York.

Fudali-Czyż, A., Francuz, P., & Augustynowicz, P. (2018). The effect of art expertise on eye fixation-related potentials during aesthetic judgment task in focal and ambient modes. *Frontiers in Psychology*, *9*, 1972.

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, *5*(1_suppl), 123–147. https://doi.org/10.1177/10298649020050S105

Glăveanu, V. P. (2014). Revisiting the "Art Bias" in lay conceptions of creativity. *Creativity Research Journal*, *26*(1), 11–20. https://doi.org/10.1080/10400419.2014.873656

Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition*, *27*, 328–338.

Hagtvedt, H., Patrick, V. M., & Hagtvedt, R. (2008). The perception and evaluation of visual art. *Empirical Studies of the Arts*, *26*(2), 197–218. https://doi.org/10.2190/EM.26.2.d

Hanich, J., Wagner, V., Shah, M., Jacobsen, T., & Menninghaus, W. (2014). Why we like to watch sad films. The pleasure of being moved in aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(2), 130–143.

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, *64*(1), 19–54.

Hitsuwari, J., & Nomura, M. (2022a). Ambiguity and beauty: Japanese-German cross-cultural comparisons on aesthetic evaluation of haiku poetry. *Psychology of Aesthetics, Creativity, and the Arts*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/aca0000497

Hitsuwari, J., & Nomura, M. (2022b). How individual states and traits predict aesthetic appreciation of haiku poetry. *Empirical Studies of the Arts*, *40*(1), 81–99. https://doi.org/10.1177/0276237420986420

Hitsuwari, J., & Nomura, M. (2022c). Ambiguity tolerance can improve through poetry appreciation and creation. *The Journal of Creative Behavior*, *n/a*(n/a). https://doi.org/https://doi.org/10.1002/jocb.574

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*(1), 33–72. https://doi.org/https://doi.org/10.1016/j.cognition.2005.11.005

Jacobs, A. M. (2017). Quantifying the beauty of words: A neurocognitive poetics perspective. In *Frontiers in Human Neuroscience* (Vol. 11). https://www.frontiersin.org/articles/10.3389/fnhum.2017.00622

Johnson-Laird, P. N., & Oatley, K. (2022). How poetry evokes emotions. *Acta Psychologica*, *224*, 103506. https://doi.org/https://doi.org/10.1016/j.actpsy.2022.103506

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575. https://doi.org/DOI: 10.1017/S0140525X08005293

Kaufman, J.C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the Consensual Assessment Technique. *The Journal of Creative Behavior*, *43*(4), 223–233. https://doi.org/https://doi.org/10.1002/j.2162-6057.2009.tb01316.x

Kaufman, James C, Baer, J., Cole, J. C., & Sexton∗, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, *20*(2), 171–178. https://doi.org/10.1080/10400410802059929

Kaufman, James C, Niu, W., Sexton, J. D., & Cole, J. C. (2010). In the eye of the beholder: differences across ethnicity and gender in evaluating creative work. *Journal of Applied Social Psychology*, *40*(2), 496–511. https://doi.org/https://doi.org/10.1111/j.1559-1816.2009.00584.x

Kenett, Y. N. (2019). What can quantitative measures of semantic distance tell us about creativity? *Current Opinion in Behavioral Sciences*, *27*, 11–16. https://doi.org/https://doi.org/10.1016/j.cobeha.2018.08.010

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews*

Neuroscience, *15*(3), 170–180. https://doi.org/10.1038/nrn3666

Kottlow, M., Praeg, E., Luethy, C., & Jancke, L. (2011). Artists' advance: decreased upper alpha power while drawing in artists compared with non-artists. *Brain Topography*, *23*, 392–402.

Kraxenberger, M., & Menninghaus, W. (2017). Affinity for poetry and aesthetic appreciation of joyful and sad poems. In *Frontiers in Psychology* (Vol. 7). https://www.frontiersin.org/articles/10.3389/fpsyg.2016.02051

Lau, J. H., Cohn, T., Baldwin, T., Brooke, J., & Hammond, A. (2018). *Deep-speare: A joint neural model of poetic language, meter and rhyme*. https://doi.org/10.48550/arxiv.1807.03491

Leder, H., Gerger, G., Dressler, S. G., & Schabmann, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(1), 2.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489–508. https://doi.org/https://doi.org/10.1348/0007126042369811

Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition and Emotion*, *28*(6), 1137–1147. https://doi.org/10.1080/02699931.2013.870132

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode – Developments and challenges in empirical aesthetics. *British Journal of Psychology*, *105*(4), 443–464. https://doi.org/https://doi.org/10.1111/bjop.12084

Lindell, A. K., & Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology*, *23*(4), 453–475. https://doi.org/10.1080/20445911.2011.539556

Liu, S., Erkkinen, M. G., Healey, M. L., Xu, Y., Swett, K. E., Chow, H. M., & Braun, A. R. (2015). Brain activity and connectivity during poetry composition: Toward a multidimensional model of the creative process. *Human Brain Mapping*, *36*(9), 3351–3372. https://doi.org/https://doi.org/10.1002/hbm.22849

Marin, M. M., & Bhattacharya, J. (2010). Music induced emotions: some current issues and cross-modal comparisons. In J. Hermida & M. Ferrero (Eds.), *Music education* (pp. 1–38). Hauppauge, NY: Nova Science Publishers.

Mastria, S., Agnoli, S., & Corazza, G. E. (2019). How does emotion influence the creativity evaluation of exogenous alternative ideas? *PLOS ONE*, *14*(7), e0219298. https://doi.org/10.1371/journal.pone.0219298

Meyer, W. U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*(3), 251–274. https://doi.org/https://doi.org/10.1023/A:1024422330338

Meyer, L. B. (1967). *Music, the arts, and ideas.*

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, *34*(8), 1388–1429. https://doi.org/https://doi.org/10.1111/j.1551-6709.2010.01106.x

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed- effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion*, *27*(7), 1326–1334. https://doi.org/10.1080/02699931.2013.777660

Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. In *Frontiers in Psychology* (Vol. 4).

https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00010

Organisciak, P., & Dumas, D. (2020). Open creativity scoring [Computer software]. *Denver, CO: University of Denver*.

Papp-Zipernovszky, O., Mangen, A., Jacobs, A., & Lüdtke, J. (2021). Shakespeare sonnet reading: An empirical study of emotional responses. *Language and Literature*, *31*(3), 296–324. https://doi.org/10.1177/09639470211054647

Plucker, J. A., Kaufman, J. C., Temple, J. S., & Qian, M. (2009). Do experts and novices evaluate movies the same way? *Psychology & Marketing*, *26*(5), 470–478. https://doi.org/https://doi.org/10.1002/mar.20283

Rasse, C., Onysko, A., & Citron, F. M. M. (2020). Conceptual metaphors in poetry interpretation: A psycholinguistic approach. *Language and Cognition*, *12*(2), 310–342. https://doi.org/10.1017/langcog.2019.47

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45–48.

Rockwood, N. J. (2017). *Advancing the formulation and testing of multilevel mediation and moderated mediation models* [The Ohio State University]. http://rave.ohiolink.edu/etdc/view?acc_num=osu1489578419777238

Scharinger, M., Wagner, V., Knoop, C. A., & Menninghaus, W. (2022). Melody in poems and songs: Fundamental statistical properties predict aesthetic evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/aca0000465

Silvia, P. J. (2009). Looking past pleasure: anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(1), 48.

Simonton, D. K. (2012). Taking the U.S. patent office criteria seriously: A quantitative three-

criterion creativity definition and its implications. *Creativity Research Journal*, *24*(2–3), 97–106. https://doi.org/10.1080/10400419.2012.676974

Warren, N. L., Farmer, M., Gu, T., & Warren, C. (2021). Marketing ideas: How to write research articles that readers understand and cite. *Journal of Marketing*, *85*(5), 42–57. https://doi.org/10.1177/00222429211003560

Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T., & Menninghaus, W. (2017). The emotional power of poetry: neural circuitry, psychophysiology and compositional principles. *Social Cognitive and Affective Neuroscience*, *12*(8), 1229–1240. https://doi.org/10.1093/scan/nsx069

Winston, A. S., & Cupchik, G. C. (1992). The Evaluation of high art and popular art by naive and experienced viewers. *Visual Arts Research*, *18*(1), 1–14. http://www.jstor.org/stable/20715763

Zioga, I., Harrison, P. M. C., Pearce, M. T., Bhattacharya, J., & Di Bernardi Luft, C. (2020). From learning to creativity: Identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity. *NeuroImage*, *206*, 116311. https://doi.org/https://doi.org/10.1016/j.neuroimage.2019.116311

Table 1

Structural and Contextual Details of the Poems

Poem Title	No. of lines	No. of words	Form	Style	Genre	Theme
Funeral Blues	16	136	Quatrain	Elegiac	Modernist	Grief
At the Same Time	10	39	Free verse	Reflective	Contemporary	Transience of existence
The Supreme Moment	16	63	Free verse	Lyrical	Contemporary	Powerlessness of humans
Peace	8	26	Quatrain	Direct	War poetry	Futility of war
A Peck of Gold	12	86	Quatrain	Lyrical	Modernist	Illusion
The Freedom of the Moon	12	91	Short lyric	Lyrical	Contemporary	Awe inspiring nature
Stopping by Woods on a Snowy Evening	16	108	Quatrain	Lyrical	Modernist	Contemplation
When you Come to me	10	41	Free verse	Introspective	Contemporary	Nostalgia
Sonnet 116	14	109	Sonnet	Formal	Romantic	Eternity of love
Apparently with no Surprise	8	36	Ballad	Narrative	Nature poetry	Transience of life
Will there really be a "Morning"?	12	76	Lyric	Playful	Lyric	Wonder and curiosity
Unable are the Loved to die	6	26	Tercet	Lyrical	Lyric	Immortality of love
How happy is the little Stone	10	46	Quatrain	Lyrical	Nature	Independence and contentment
Ah! Sun-flower	8	50	Quatrain	Lyrical	Romantic	Individualism

The Smile	16	106	Quatrain	Lyrical	Romantic	Love and deceit
Love after Love	15	98	Free verse	Introspective	Contemporary	Self-love, self-discovery
The Last Laugh	15	90	Free verse	onomatopoeic	War poem	Brutality of war
Clown in the Moon	8	48	Quatrain	Lyrical	Lyric	Melancholy
Remember	14	111	Petrarchan Sonnet	Lyrical	Lyric	Love, death, remembrance
When You Are Old	12	100	Short lyric	Romantic	Romantic	True love
Aedh wishes for the Cloths of Heaven	9	60	Free verse	Romantic	Romantic	Love and dreams
Memory	6	33	Free verse	Lyrical	Lyric	Transience of youth and beauty
The Arrow and the Song	12	88	Free verse	Narrative	Lyric	Power of words and friendship
There Will Come Soft Rains	12	91	Lyric	Lyrical	Lyric	War
Alone	12	80	Quatrain	Romantic	Romantic	Solitude
Ozymandias	14	111	Sonnet	Romantic	Romantic	Transience of power
Good-Night	12	75	Quatrain	Romantic	Romantic	Longing for togetherness
A Lament	10	65	Short lyric	Ethereal	Lyric	Nostalgia, sadness
Alas! This Is Not What I Thought Life Was.	9	75	Short lyric	Introspective	Lyric	Challenges of life
Fragment: A Wanderer	4	25	Quatrain	Romantic	Romantic	Imagination
Fragment: Apostrophe to Silence	9	69	Short lyric	Introspective	Romantic	Existential longing

MEDITATION XVII	13	81	Free verse	Metaphysical	Romantic	Interconnectedness of humanity
Nothing Gold Can Stay	8	40	Octave	Lyrical	Nature	Impermanence of beauty
My Heart Leaps Up	9	61	Quatrain	Lyrical	Nature	Ecstasy with nature’s beauty
I taste a liquor never brewed –	16	81	Lyric	Playful	Lyric	Joy of Nature
If You Should Go	8	44	Quatrain	Lyrical	Romantic	Transience of love and life

Table 2

Descriptive Statistics of the Variables, Including Means (M), Standard Deviation (SD), Skewness, Kurtosis, Standard Error (SE) and Variance Inflation Factor (VIF)

Variable	*n*	*mean*	*sd*	*median*	*min*	*max*	*skew*	*kurtosis*	*se*	*VIF*
Clarity	3456	4.82	1.58	5.0	1	7	-0.46	-0.57	0.03	1.58
Aesthetic Appeal	3456	4.80	1.44	5.0	1	7	-0.48	-0.23	0.02	2.13
Felt Valence	3456	4.50	1.62	5.0	1	7	-0.41	-0.48	0.03	2.59
Felt arousal	3456	3.86	1.73	4.0	1	7	-0.14	-0.92	0.03	2.00
Surprise	3456	3.78	1.68	4.0	1	7	-0.17	-0.92	0.03	1.63
Creativity	3456	4.91	1.38	5.0	1	7	-0.53	0.05	0.02	

Table 3

Bivariate Correlations Among Variables Showing the Strongest Correlation between Aesthetic Appeal and Creativity

Variable	*M*	*SD*	1	2	3	4	5
1. Clarity	4.82	1.58	-				
2. Aesthetic Appeal	4.80	1.44	.57**	-			
3. Felt Valence	4.50	1.62	.54**	.68**	-		
4. Felt Arousal	3.86	1.73	.38**	.51**	.64**	-	
5. Surprise	3.78	1.68	.31**	.43**	.53**	.59**	-
6. Creativity	4.91	1.38	.42**	.66**	.60**	.47**	.52**

Note. M and *SD* are used to represent mean and standard deviation, respectively. ** indicates $p < .01$.

Table 4

Comparison of Models Using Model-Fit Criteria Indicating that the Most Parsimonious Model Comprises Aesthetic Appeal, Felt Valence, and Surprise

Model fit criteria	Null Model	Model 1	Model 2	Model3	Model 4	Model 5
AIC	11160.3	9586.0	9305.5	9020.0	9000.8	8999.5
BIC	11178.7	9622.8	9354.7	9081.5	9074.6	9085.5
R^2(m)	.00	.27	.30	.34	.34	.34
$\Delta\chi^2$(df)		1580.3***	284.5***	289.5***	23.1***	5.3

Note. Aesthetic appeal, felt valence, surprise, arousal and clarity are included sequentially to Model 1 to Model 5; all models are compared hierarchically, i.e., Model 1 is compared to Null Model, Model 2 is compared to Model 1 and so on ; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; R2 (m) = proportion of variation explained by fixed effects (Nakagawa & Schielzeth, 2013); $\Delta\chi 2$ = Likelihood ratio test statistic for comparison of models. *** $p < .001$.

Table 5

Linear Mixed Model Results for the Best Model-Fit Comprising Aesthetic Appeal, Surprise and Felt Valence as Potential Predictors of Creativity Judgment of Poems

Random effects:

Groups	Name	Variance	Std. Dev.
Participants	(Intercept)	0.54	0.74
Residual		0.72	0.85

Number of observations: 3456, groups: Participants, 96

Fixed effects:

	Estimate	S.E.	df	*t*-value	*p*-value
(Intercept)	5.01	0.10	96.00	50.43	0.00
Expertise	-0.25	0.16	96.00	-1.58	0.12
Aesthetic appeal	0.31	0.02	3360.00	15.16	0.00
Felt valence	0.20	0.02	3360.00	10.27	0.00
Surprise	0.25	0.02	3360.00	14.72	0.00
Expertise*Aesthetic appeal	0.07	0.03	3360.00	2.33	0.02
Expertise*Felt valence	-0.08	0.03	3360.00	-2.76	0.01
Expertise*Surprise	-0.06	0.03	3360.00	-2.37	0.02

Grouping Variables

Group	*# groups*	*ICC*
Participants	*96*	*0.43*

Table 6

The Results of the Simple Slopes Analyses for Expertise-Moderated Creativity Judgment Considering Participants (N=39) with Formal Degree in English Literature as Experts

	Nonexperts (0)				Experts (1)				Slope difference Nonexpert– Expert		
Variable	*b*	*SE*	95% CI		*b*	*SE*	95% CI		*b*	*z-ratio*	*p-value*
			LCL	*UCL*			*LCL*	*UCL*			
AA	.31	.02	.27	.35	.38	.02	.34	.43	-.07	-2.33	.02
FV	.20	.02	.16	.24	.12	.02	.08	.16	.08	2.80	.01
Surp	.25	.02	.22	.29	.19	.02	.15	.23	.06	2.40	.02

Note. Degrees-of-freedom method: asymptotic; Confidence level used: 0.95; Exp = Expertise, AA=Aesthetic appeal, FV=Felt valence, Surp = Surprise; LCL = Lower confidence level; UCL = Upper confidence level; Confidence level used: 0.95.

Table 7

Semantic Distance Scores between Subjectively Chosen Surprise Evoking Line (s) and the Two Preceding Lines Using Two Automated Scoring Measures: SemDis and Open Creativity Scoring (OCS)

	SemDis				OCS			
Measure	*b*	*SE*	*t*-val	*p*-val	*b*	*SE*	*t*-val	*p*-val
Intercept	3.78	0.11	33.22	< .001	3.78	0.11	33.22	< .001
Scores	1.22	0.07	17.43	< .001	2.18	0.10	22.28	< .001

Figure 1

Parallel Multilevel Mediation with Felt Valence and Surprise as Mediators on the Relationship Between Aesthetic Appeal and Creativity Demonstrates Partial Mediations by Both Mediators

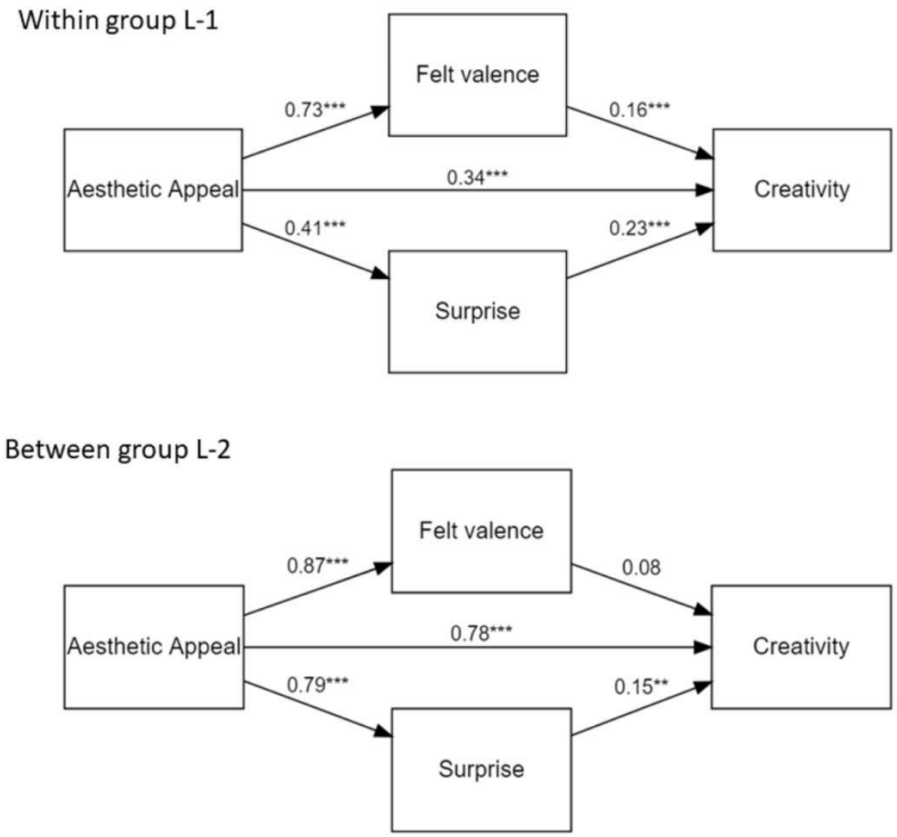

Figure 2

The Expertise Moderated Creativity Evaluation Shows Significant Moderation by Expertise on Aesthetic Appeal, Felt Valence and Surprise in Judging Creativity of Poem

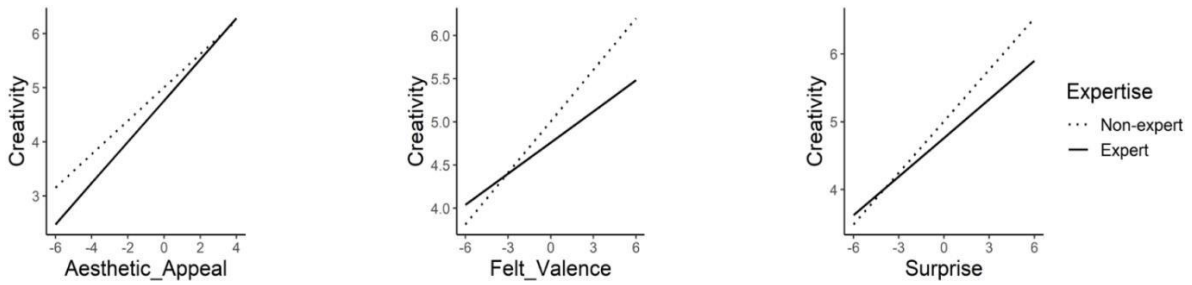

Table S1

The Results of the Simple Slopes Analyses for Expertise-Moderated Creativity Judgment Considering Participants (N= 21) with Long Association (>=10 Years) with English Poetry as Experts

	Nonexperts (0)				Experts (1)				Slope difference Nonexpert– Expert		
Variable	*b*	*SE*	95% CI		*b*	*SE*	95% CI		*b*	*z-ratio*	*p-value*
			LCL	*UCL*			*LCL*	*UCL*			
AA	.34	.02	.31	.38	.33	.03	.26	.40	.01	.34	.74
FV	.16	.02	.13	.19	.19	.04	.12	.26	-.04	-.92	.36
Surp	.24	.01	.21	.26	.21	.03	.15	.27	.02	.73	.47

Note. Degrees-of-freedom method: asymptotic; Confidence level used: 0.95. Exp = Expertise, AA = Aesthetic appeal, FV = Felt valence, Surp = Surprise; LCL = Lower confidence level; UCL = Upper confidence level; Confidence level used: 0.95.

Table S2

Comparison of Moderation Results with Two Different Criteria of Expertise Indicates that Expertise in English Literature Significantly Moderates Aesthetic Appeal, Felt Valence and Surprise whereas Long Association with Poetry (>= 10 years) Does Not Moderate Any of the Predictors

	Expertise criteria							
	Formal degree in English(*N*=39)				>= 10 years of association(*N*=21)			
Model	*b*	*SE*	*t*-val	*p*-val	*b*	*SE*	*t*-val	*p*-val
Exp*AesthApp	.07	.03	2.33	.02	-.01	.04	.34	.74
Exp*FeltValence	-.08	.03	-2.76	.01	.04	.04	-.92	.36
Exp*Surprise	-.06	.03	-2.37	.02	-.02	.03	-.73	.47

Note. AesthApp=Aesthetic Appeal

Table S3

Moderation Results (N=126) with Two Different Criteria of Expertise Indicate that Expertise in English Literature Significantly Moderates Aesthetic Appeal, Felt Valence and Surprise and Long Association with Poetry (>=10 years) Moderates Aesthetic Appeal and Surprise

	Expertise criteria (*N* =126)							
Moderation	Formal degree in English (*N* = 49)				>= 10 years of association (*N* =32)			
	b	*SE*	*t*-val	*p*-val	*b*	*SE*	*t*-val	*p*-val
Exp*AesthApp	.06	.03	2.22	.03	.07	.03	2.16	.03
Exp*FeltValence	-.05	.02	-2.19	.03	.04	.03	1.17	.24
Exp*Surprise	-.06	.02	-3.46	<.001	-.06	.03	-2.04	.04

Note. AesthApp = Aesthetic Appeal

List of Tables and Figures

Figures

Supplementary materials

Figure 1. Parallel Multilevel Mediation with Felt Valence and Surprise as Mediators on the Relationship Between Aesthetic Appeal and Creativity Demonstrates Partial

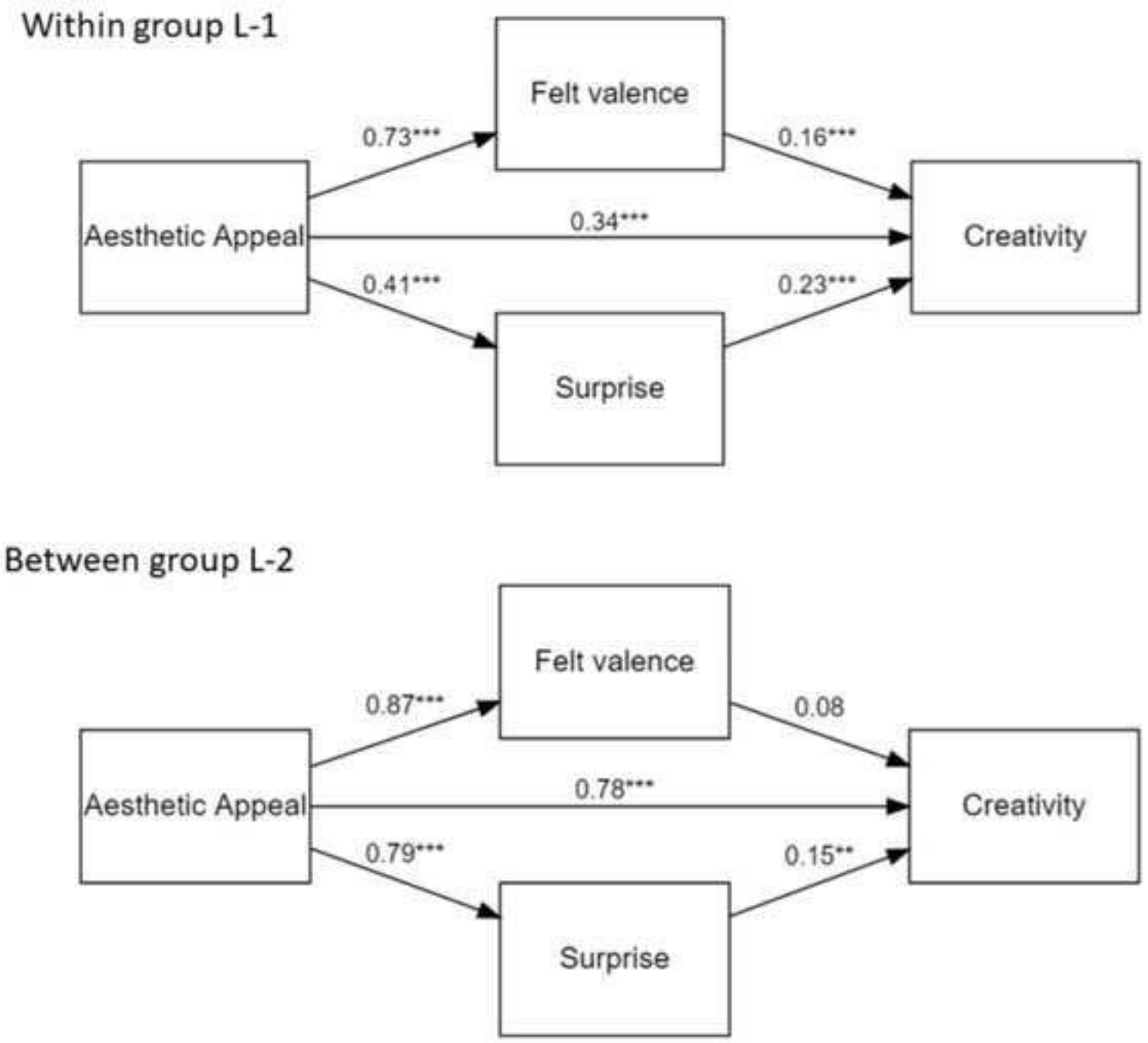

Figure 2. The Expertise Moderated Creativity Evaluation Shows Significant Moderation by Expertise on Aesthetic Appeal, Felt Valence, and Surprise in Judging Creativity of

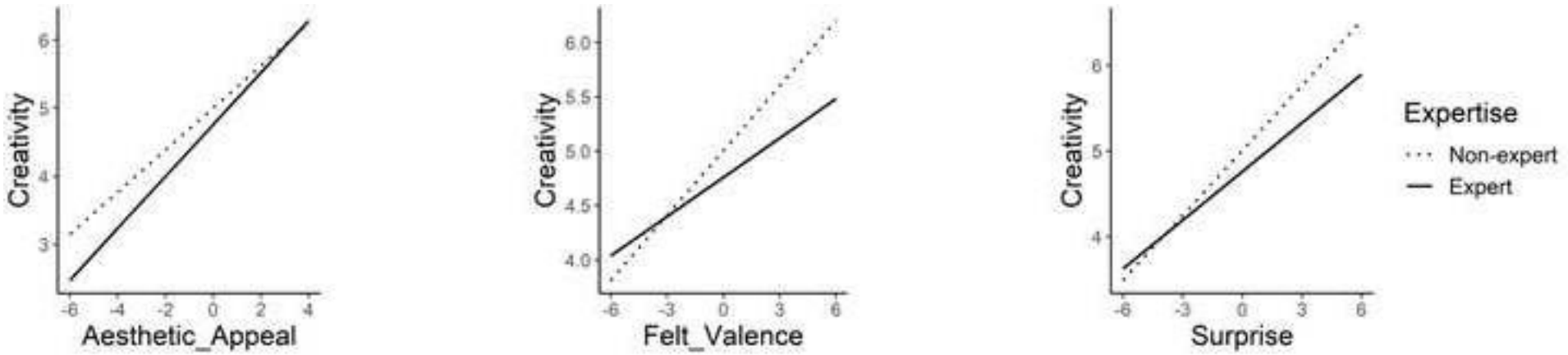

Table 1

Structural and Contextual Details of the Poems

Poem Title	No. of lines	No. of words	Form	Style	Genre	Theme
Funeral Blues	16	136	Quatrain	Elegiac	Modernist	Grief
At the Same Time	10	39	Free verse	Reflective	Contemporary	Transience of existence
The Supreme Moment	16	63	Free verse	Lyrical	Contemporary	Powerlessness of humans
Peace	8	26	Quatrain	Direct	War poetry	Futility of war
A Peck of Gold	12	86	Quatrain	Lyrical	Modernist	Illusion
The Freedom of the Moon	12	91	Short lyric	Lyrical	Contemporary	Awe inspiring nature
Stopping by Woods on a Snowy Evening	16	108	Quatrain	Lyrical	Modernist	Contemplation
When you Come to me	10	41	Free verse	Introspective	Contemporary	Nostalgia
Sonnet 116	14	109	Sonnet	Formal	Romantic	Eternity of love
Apparently with no Surprise	8	36	Ballad	Narrative	Nature poetry	Transience of life
Will there really be a "Morning"?	12	76	Lyric	Playful	Lyric	Wonder and curiosity
Unable are the Loved to die	6	26	Tercet	Lyrical	Lyric	Immortality of love
How happy is the little Stone	10	46	Quatrain	Lyrical	Nature	Independence and contentment
Ah! Sun-flower	8	50	Quatrain	Lyrical	Romantic	Individualism

The Smile	16	106	Quatrain	Lyrical	Romantic	Love and deceit
Love after Love	15	98	Free verse	Introspective	Contemporary	Self-love, self-discovery
The Last Laugh	15	90	Free verse	onomatopoeic	War poem	Brutality of war
Clown in the Moon	8	48	Quatrain	Lyrical	Lyric	Melancholy
Remember	14	111	Petrarchan Sonnet	Lyrical	Lyric	Love, death, remembrance
When You Are Old	12	100	Short lyric	Romantic	Romantic	True love
Aedh wishes for the Cloths of Heaven	9	60	Free verse	Romantic	Romantic	Love and dreams
Memory	6	33	Free verse	Lyrical	Lyric	Transience of youth and beauty
The Arrow and the Song	12	88	Free verse	Narrative	Lyric	Power of words and friendship
There Will Come Soft Rains	12	91	Lyric	Lyrical	Lyric	War
Alone	12	80	Quatrain	Romantic	Romantic	Solitude
Ozymandias	14	111	Sonnet	Romantic	Romantic	Transience of power
Good-Night	12	75	Quatrain	Romantic	Romantic	Longing for togetherness
A Lament	10	65	Short lyric	Ethereal	Lyric	Nostalgia, sadness
Alas! This Is Not What I Thought Life Was.	9	75	Short lyric	Introspective	Lyric	Challenges of life
Fragment: A Wanderer	4	25	Quatrain	Romantic	Romantic	Imagination
Fragment: Apostrophe to Silence	9	69	Short lyric	Introspective	Romantic	Existential longing

MEDITATION XVII	13	81	Free verse	Metaphysical	Romantic	Interconnectedness of humanity
Nothing Gold Can Stay	8	40	Octave	Lyrical	Nature	Impermanence of beauty
My Heart Leaps Up	9	61	Quatrain	Lyrical	Nature	Ecstasy with nature’s beauty
I taste a liquor never brewed –	16	81	Lyric	Playful	Lyric	Joy of Nature
If You Should Go	8	44	Quatrain	Lyrical	Romantic	Transience of love and life

Table 2

Descriptive Statistics of the Variables, Including Means (M), Standard Deviation (SD), Skewness, Kurtosis, Standard Error (SE) and Variance Inflation Factor (VIF)

Variable	*n*	*mean*	*sd*	*median*	*min*	*max*	*skew*	*kurtosis*	*se*	*VIF*
Clarity	3456	4.82	1.58	5.0	1	7	-0.46	-0.57	0.03	1.58
Aesthetic Appeal	3456	4.80	1.44	5.0	1	7	-0.48	-0.23	0.02	2.13
Felt Valence	3456	4.50	1.62	5.0	1	7	-0.41	-0.48	0.03	2.59
Felt arousal	3456	3.86	1.73	4.0	1	7	-0.14	-0.92	0.03	2.00
Surprise	3456	3.78	1.68	4.0	1	7	-0.17	-0.92	0.03	1.63
Creativity	3456	4.91	1.38	5.0	1	7	-0.53	0.05	0.02	

Table 3

Bivariate Correlations Among Variables Showing the Strongest Correlation between Aesthetic Appeal and Creativity

Variable	*M*	*SD*	1	2	3	4	5
1. Clarity	4.82	1.58	-				
2. Aesthetic Appeal	4.80	1.44	.57**	-			
3. Felt Valence	4.50	1.62	.54**	.68**	-		
4. Felt Arousal	3.86	1.73	.38**	.51**	.64**	-	
5. Surprise	3.78	1.68	.31**	.43**	.53**	.59**	-
6. Creativity	4.91	1.38	.42**	.66**	.60**	.47**	.52**

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. ** indicates $p < .01$.

Table 4

Comparison of Models Using Model-Fit Criteria Indicating that the Most Parsimonious Model Comprises Aesthetic Appeal, Felt Valence, and Surprise

Model fit criteria	Null Model	Model 1	Model 2	Model3	Model 4	Model 5
AIC	11160.3	9586.0	9305.5	9020.0	9000.8	8999.5
BIC	11178.7	9622.8	9354.7	9081.5	9074.6	9085.5
R^2(m)	.00	.27	.30	.34	.34	.34
$\Delta\chi^2$(df)		1580.3***	284.5***	289.5***	23.1***	5.3

Note. Aesthetic appeal, felt valence, surprise, arousal and clarity are included sequentially to Model 1 to Model 5; all models are compared hierarchically, i.e., Model 1 is compared to Null Model, Model 2 is compared to Model 1 and so on ; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; R2 (m) = proportion of variation explained by fixed effects (Nakagawa & Schielzeth, 2013); $\Delta\chi 2$ = Likelihood ratio test statistic for comparison of models. *** $p < .001$.

Table 5

Linear Mixed Model Results for the Best Model-Fit Comprising Aesthetic Appeal, Surprise and Felt Valence as Potential Predictors of Creativity Judgment of Poems

Random effects:

Groups	Name	Variance	Std. Dev.
Participants	(Intercept)	0.54	0.74
Residual		0.72	0.85

Number of observations: 3456, groups: Participants, 96

Fixed effects:

	Estimate	S.E.	df	*t*-value	*p*-value
(Intercept)	5.01	0.10	96.00	50.43	0.00
Expertise	-0.25	0.16	96.00	-1.58	0.12
Aesthetic appeal	0.31	0.02	3360.00	15.16	0.00
Felt valence	0.20	0.02	3360.00	10.27	0.00
Surprise	0.25	0.02	3360.00	14.72	0.00
Expertise*Aesthetic appeal	0.07	0.03	3360.00	2.33	0.02
Expertise*Felt valence	-0.08	0.03	3360.00	-2.76	0.01
Expertise*Surprise	-0.06	0.03	3360.00	-2.37	0.02

Grouping Variables		
Group	*# groups*	*ICC*
Participants	*96*	*0.43*

Table 6

The Results of the Simple Slopes Analyses for Expertise-Moderated Creativity Judgment Considering Participants (N=39) with Formal Degree in English Literature as Experts

	Nonexperts (0)				Experts (1)				Slope difference Nonexpert– Expert		
Variable	*b*	*SE*	95% CI		*b*	*SE*	95% CI		*b*	*z-ratio*	*p-value*
			LCL	*UCL*			*LCL*	*UCL*			
AA	.31	.02	.27	.35	.38	.02	.34	.43	-.07	-2.33	.02
FV	.20	.02	.16	.24	.12	.02	.08	.16	.08	2.80	.01
Surp	.25	.02	.22	.29	.19	.02	.15	.23	.06	2.40	.02

Note. Degrees-of-freedom method: asymptotic; Confidence level used: 0.95; Exp = Expertise, AA=Aesthetic appeal, FV=Felt valence, Surp = Surprise; LCL = Lower confidence level; UCL = Upper confidence level; Confidence level used: 0.95.

Table 7

Semantic Distance Scores between Subjectively Chosen Surprise Evoking Line (s) and the Two Preceding Lines Using Two Automated Scoring Measures: SemDis and Open Creativity Scoring (OCS)

	SemDis				OCS			
Measure	*b*	*SE*	*t*-val	*p*-val	*b*	*SE*	*t*-val	*p*-val
Intercept	3.78	0.11	33.22	< .001	3.78	0.11	33.22	< .001
Scores	1.22	0.07	17.43	< .001	2.18	0.10	22.28	< .001

Table S1

The Results of the Simple Slopes Analyses for Expertise-Moderated Creativity Judgment Considering Participants (N= 21) with Long Association (>=10 Years) with English Poetry as Experts

	Nonexperts (0)				Experts (1)				Slope difference Nonexpert– Expert		
Variable	*b*	*SE*	95% CI		*b*	*SE*	95% CI		*b*	*z-ratio*	*p-value*
			LCL	*UCL*			*LCL*	*UCL*			
AA	.34	.02	.31	.38	.33	.03	.26	.40	.01	.34	.74
FV	.16	.02	.13	.19	.19	.04	.12	.26	-.04	-.92	.36
Surp	.24	.01	.21	.26	.21	.03	.15	.27	.02	.73	.47

Note. Degrees-of-freedom method: asymptotic; Confidence level used: 0.95. Exp = Expertise, AA = Aesthetic appeal, FV = Felt valence, Surp = Surprise; LCL = Lower confidence level; UCL = Upper confidence level; Confidence level used: 0.95.

Table S2

Comparison of Moderation Results with Two Different Criteria of Expertise Indicates that Expertise in English Literature Significantly Moderates Aesthetic Appeal, Felt Valence and Surprise whereas Long Association with Poetry (>= 10 years) Does Not Moderate Any of the Predictors

	Expertise criteria							
	Formal degree in English(*N*=39)				>= 10 years of association(*N*=21)			
Model	*b*	*SE*	*t*-val	*p*-val	*b*	*SE*	*t*-val	*p*-val
Exp*AesthApp	.07	.03	2.33	.02	-.01	.04	.34	.74
Exp*FeltValence	-.08	.03	-2.76	.01	.04	.04	-.92	.36
Exp*Surprise	-.06	.03	-2.37	.02	-.02	.03	-.73	.47

Note. AesthApp=Aesthetic Appeal

Table S3

Moderation Results (N=126) with Two Different Criteria of Expertise Indicate that Expertise in English Literature Significantly Moderates Aesthetic Appeal, Felt Valence and Surprise and Long Association with Poetry (>=10 years) Moderates Aesthetic Appeal and Surprise

	Expertise criteria (*N* =126)							
Moderation	Formal degree in English (*N* = 49)				>= 10 years of association (*N* =32)			
	b	*SE*	*t*-val	*p*-val	*b*	*SE*	*t*-val	*p*-val
Exp*AesthApp	.06	.03	2.22	.03	.07	.03	2.16	.03
Exp*FeltValence	-.05	.02	-2.19	.03	.04	.03	1.17	.24
Exp*Surprise	-.06	.02	-3.46	<.001	-.06	.03	-2.04	.04

Note. AesthApp = Aesthetic Appeal