The Aesthetic Responsiveness Assessment (AReA): A screening tool to assess individual differences in responsiveness to art in English and German

Wolff Schlotz* [1, 2], Sebastian Wallot* [1], Diana Omigie[1, 3], Michael D. Masucci[4], Sonja C. Hoelzmann[1, 2], and Edward A. Vessel[1]

[1] Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

[2] Insitute of Psychology, Gothe-University Frankfurt am Main, Germany

[3] Department of Psychology, Goldsmiths, University of London, UK

[4] Department of Psychology, New York University

Author Note

* These authors contributed equally to this article and hence share first authorship

Sebastian Wallot, https://orcid.org/0000-0002-3626-3940

Wolff Schlotz, https://orcid.org/0000-0003-2356-7766

Diana Omigie, https://orcid.org/0000-0003-1600-0112

Michael D. Masucci, https://orcid.org/0000-0002-9151-7253

Sonja C. Hoelzmann, https://orcid.org/0000-0002-7655-5884

Edward A. Vessel, https://orcid.org/0000-0001-8733-1731

Michael D. Masucci is now at Louisiana State University, Louisiana

We have no known conflict of interest to disclose.

The authors would like to acknowledge Amy Belfi, Anna Kasdan, Gabrielle Starr and Jonathan Stahl for help with data collection, Christine Knoop for substantially contributing to the translation of the AReA into German, and Kirill Fayn for commenting on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Sebastian Wallot, Max Planck Institute for Empirical Aesthetics, Grueneburgweg 14, 60322 Frankfurt am Main, Germany. E-mail: sebastian.wallot@ae.mpg.de

Abstract

People differ in how they respond to artworks. Measuring such individual differences is helpful for explaining response variability and selecting particularly responsive sub-samples. On the basis of a sample of items indicating relevant behavior and experience, we exploratively constructed the Aesthetic Responsiveness Assessment (AReA), a screening tool for the assessment of individual differences in responsiveness to art in English and German. Exploratory and confirmatory factor analyses suggested three first-order factors labeled aesthetic appreciation, intense aesthetic experience, and creative behavior, and a second-order factor aesthetic responsiveness. Aesthetic responsiveness was assessed in $N = 781$ participants from the United States and Germany, and measurement invariance analysis demonstrated full metric and partial scalar invariance across language versions. AReA scale scores yielded good reliability estimates. Validation studies confirmed expected associations between AReA scale scores and measures of related constructs, as well as continuously and retrospectively recorded responses to music, visual art, and poetry. In summary, the AReA is a promising, psychometrically evaluated instrument to assess aesthetic responsiveness built on a mixture of exploratory and confirmatory construction strategies. It can be used as a screening tool both in English and German speaking samples.

Keywords: aesthetic responsiveness, creative behavior, aesthetic experience, screening scale, validity, measurement invariance

The Aesthetic Responsiveness Assessment (AReA): A screening tool to assess individual differences in responsiveness to art in English and German

There exist individual differences in responsiveness to many different types of information (e.g. to visual brightness, auditory loudness, taste, social or emotional cues), and responsiveness to aesthetic stimuli is no exception. Indeed, aesthetic experiences would appear to be a domain where individual differences in responsiveness are rather large. We may all call to mind individuals whose responsiveness is different than our own: for instance, a colleague may report that they generally don't get pleasure from visiting museums, or from listening to music. In contrast, we may know other individuals whose level of aesthetic responsiveness to a particular art form is so strong as to be wholly out of our level of understanding.

As experimentalists interested in studying the psychological and neural basis of aesthetic experiences, this heterogeneity in aesthetic responsiveness presents a distinct problem. If a large proportion of the potential observers that we sample from the general population do not respond to our stimuli, this may result in inconclusive findings. While at least a portion of variability may reflect individual preferences for specific aesthetic domains or styles, part of this variability likely also reflects trait-level differences in overall aesthetic responsiveness. Here, we present a screening tool developed with the goal of providing a quick assessment of (overall) aesthetic responsiveness.

We define aesthetic responsiveness here as the individual capacity to respond to aesthetic stimuli. This definition is mainly based on the notion that aesthetic responses have a common origin in brain areas that mediate responses across different domains, particularly neural systems involved in emotion and reward processing (Berlyne, 1971; Chatterjee & Vartanian, 2016; Vessel et al., 2019). These neural systems can affect peripheral responses via connections with the autonomic nervous and neuroendocrine systems that link central

nervous system activity with peripheral physiological responses (Lane et al., 2009). This conceptualization of aesthetic responsiveness implies some sort of generality, such that individual differences in responsiveness may exist across aesthetic domains, response domains (cognitive, emotional, behavioral, and physiological), and time (e.g., repeated exposure). However, this does not rule out stimulus specificity whereby aesthetic stimuli of different domains may result in systematically different aesthetic experiences, for example due to perceptual modality-dependent processing (cf. Jacobsen & Beudt, 2017). In addition, we acknowledge here that some response variance is likely to be due to individual-specific responses, i.e. patterns of responses that differ systematically between individuals (Vessel et al., 2018).

We assume that aesthetic responsiveness is a dispositional tendency that generates individual differences in responses to aesthetic stimuli. These individual differences are assumed to be relatively consistent over time and across aesthetic domains, as well as coherent across response domains. It is assumed that individuals with a high aesthetic responsiveness trait level experience aesthetic cognition, emotion and related physiological effects more frequently and more intensively than others, and that they show a greater behavioral propensity towards engagement with art.

The construct of aesthetic responsiveness is related to constructs focusing on individual differences in the appreciation of, or engagement with beauty (Diessner et al., 2018; Diessner et al., 2008; Haidt & Keltner, 2004), particularly if appreciation is conceived as a cognitive-emotional, and engagement as an emotional reaction to beauty (Güsewell & Ruch, 2012). However, aesthetic responsiveness differs from these constructs in a number of aspects. First, it focuses on responses to aesthetic stimuli and excludes non-aesthetic stimuli such as talent, virtue, or morality. Second, it explicitely distinguishes between response domains, providing a background for more fine-grained predictions of domain-specific

responses. Finally, aesthetic responsiveness does not exclusively focus on beauty; it includes responses to aesthetic stimuli that are not necessarily perceived as beautiful.

Regarding associations of aesthetic responsiveness with personality factors, openness to experience (or open-mindedness) seems to be particularly relevant. Findings from empirical aesthetics studies investigating openness demonstrate that personality is predictive of indicators of aesthetic experience (Fayn et al., 2015; McCrae, 2007; Rawlings et al., 2000; Silvia et al., 2015). Openness has also been linked with aesthetic activities and positive aesthetic attitudes (McManus & Furnham, 2006). Measurements of aesthetic responsiveness should therefore show strong associations with measurements of openness. In comparison to constructs of major taxonomies of personality traits, aesthetic responsiveness is closely linked, conceptually, with a specific facet related to aesthetic experience which is located in the lower level structure of the factor openness. This facet has been labelled aesthetics (Costa & McCrae, 1995), aesthetic sensitivity (Soto & John, 2017), or aesthetic appreciation (Ashton & Lee, 2007). However, openness additionally comprises a number of facets that are not part of the construct of aesthetic responsiveness. For example, a detailed analysis found five facets of openness in addition to the facet aesthetics which have been labeled intellectual efficiency, ingenuity, curiosity, tolerance, and depth (Woo et al., 2014). While these lower level facets can be expected to be empirically related to aesthetic responsiveness, they clearly reflect different constructs. Thus, while aesthetic responsiveness is thought to be similar to the openness facet aesthetics, openness is a much broader construct comprising facets that are clearly distinguishable from aesthetic responsiveness both empirically and with regard to content.

As opposed to the concept of aesthetic sensitivity, which has historically been identified as the degree to which an individuals' aesthetic judgments agree with an externally defined standard (Child, 1964; Eysenck, 1940), aesthetic responsiveness is defined by the

strength of the response, regardless of an individual's subjective sense of taste. Therefore, evaluative constructs as assessed by aesthetic sensitivity tests should be empirically distinguishable from aesthetic responsiveness as well as related constructs such as the personality factor openness. In line with this assumption, individual scores on the Visual Aesthetic Sensitivity Test (Götz et al., 1979), a measure of aesthetic sensitivity, showed only a modest correlation with the openness facet scale Aesthetics (Myszkowski et al., 2014).

As a more convenient alternative to a complete assessment of aesthetic responsiveness across all possible aesthetic domains and response domains (e.g. behavioral, physiological, emotional, cognitive), we present a self-resport assessment tool of how individuals have perceived their responses in different stimulus and response domains in their daily life. This approach is particularly useful for screening for individual aesthetic responsiveness in research settings that do not allow for rigorous and comprehensive testing that encompasses all domains.

Similar scales have been developed for different aesthetic domains, and represent different aspects of aesthetic responsiveness to a greater or lesser degree (Hager et al., 2012; Rowold, 2008; Stamatopoulou, 2004). This includes a recent scale that provides a very fine-grained assessment of aesthetic-emotional responses (Schindler et al., 2017). The measure that reflects a construct most closely related to aesthetic responsiveness is the Engagement with Beauty Scale (EBS; Diessner et al., 2008), which itself is related to the Appreciation of Beauty and Excellence (ABE) subscale of the Values in Action Inventory of Strengths (VIA-IS; Peterson & Seligman, 2004). However, the EBS focuses exclusively on the experience of beauty and is designed to measure engagement with beauty across natural, artistic, and moral domains. This wider scope is not a good match for a more focused assessment of aesthetic responsiveness. Additionally, the EBS does not separate out aesthetic responsiveness to different artistic domains, nor does it assess behavioral indicators of art appreciation. Taken

together, none of the existing instruments assesses the breadth of aesthetic responsiveness specific to artworks as defined above with a short scale that can be used for screening purposes.

We will here present rationale and choices of constructing a scale for the assessment of aesthetics responsiveness that assesses individual responses to aesthetically relevant stimuli from a broad variety of different domains. We present analyses of psychometric properties of two language versions of the scale, English and German. In the subsequent sections, we present results from a number of studies that provided data we used for validation of the scale, namely correlations of scale scores with individual responses to visual art, music, and poetry, as well as with measures of related personality constructs. Finally, a validation study will be presented, where participants filled in the resulting scale together with a measure of the Big Five personality domains and their factes; the analysis focuses on correlations of scale scores with openness and its facets.

Scale Construction

With a focus on research participant screening for aesthetic responsiveness, an 18 item short scale was developed in the English language, assessing typical responses to and engagement with a variety of aesthetic stimuli, and with an emphasis on visual aesthetic experiences to reflect that a large proportion of art has a visual component (painting, sculpture, dance, film, etc.). Due to the self-report format, the scale assesses *perceived* (self-evaluated) aesthetic responsiveness, reflecting typical and daily life aesthetic experiences. The items were designed with the aim of assessing general or aggregate experiences, in contrast to focusing on single episodes.

One goal of scale construction was to reflect the centrality of “beauty” as a core domain-general aesthetic emotion term (Istok et al., 2009; Jacobsen et al., 2004; Menninghaus et al., 2019) but also to acknowledge that this is not the only path to positive aesthetic

experiences, and that research participants often misinterpret “beauty” to refer to objective stimulus traits rather than as an emotional responding arising from the interaction of a perceiver with an object (Reber et al., 2004; Vessel et al., 2012).

Another key goal of scale construction was to distinguish between those individuals who regularly respond to artworks in an intense way from those who rarely experience more than a commonplace appreciation of aesthetic objects in everyday life. Recent empirical work suggests a potential difference between more everyday positive experiences of beauty and a subset of more intense aesthetic experiences (e.g. “being moved”, “awe”, the “sublime”; (Brielmann & Pelli, 2017; Omigie et al., 2019; Pelowski et al., 2017; Vessel et al., 2012, 2013).

Such work parallels accounts in the philosophical literature that pit feelings of beauty against those of the sublime (Burke, 1757/2015). In the context of music, for instance, beauty experiences “in which tension and discord have at most a minor place” have been distinguished from other forms of beauty, that may, instead, confront or challenge (Levinson, 2012, p. 128). Here, we sought to extend, to the individual differences level, this notion of a distinction in the types of aesthetic states that are possible. We propose that a scale that is able to reveal those individuals that regularly respond to artworks in an intense way would allow experimenters to better account for much variability in responses observable in their data.

Another goal of scale construction was to differentiate individuals who actively occupy themselves with the creation of aesthetically relevant products from those who do not. Although creative behavior does not reflect aesthetic responsiveness at the same level as appreciation of aesthetic objects does, we assume that individuals high in aesthetic responsiveness have a higher propensity to actively engage in goal-directed creative processes such as writing, painting, or making music. On the one hand, this is based on well-established associations between openness and creativity (Puryear et al., 2017), suggesting that openness

contributes substantially to an individual's creative potential. On the other hand, the link of creative potential with actual creative behavior is assumed to be moderated by a number of factors, suggesting that creative potential can or cannot lead to creative behavior (e.g. Karwowski & Beghetto, 2019). We assume that individuals high on aesthetic responsiveness have a higher creative potential, and that creative behavior is therefore linked with aesthetic responsiveness. However, this link is thought to be moderate, as other factors influence creative potential and its effect on creative behavior. We added items on creative behavior to the scale, thereby broadening the scope of the construct measurement. While emotional, cognitive, and physiological responses to aesthetic stimuli were covered by many items, behavioral indicators of aesthetic responsiveness were represented less well. Therefore, including items assessing creative behavior brings the representation of indicators of different construct-relevant responses to a similar level. While creative behavior seems to be a rather distal indicator of aesthetic responses, it should be kept in mind that it requires continued preoccupation with aesthetically relevant material and therefore reflects an individual's receptiveness for such material. The inclusion of items related to creative behaviour also aimed to achieve more precise measurements by separating variance components indicating different facets of aesthetic responsiveness. Moreover, adding creative behavior items might be particularly relevant for selecting participants for studies focusing on creative behavior, and therefore potentially increase the utility of the scale.

We began by modifying several items from the EBS reflecting experiences with artworks and expanding these into a set of eight questions reflecting either beauty or intense aesthetic experience, across four response domains: cognition (items 3, 16), physiological arousal (items 8, 10), conscious emotion (18, 13) and spirituality/transcendence (items 5, 14). Next, a set of five questions were added to assess aesthetic appreciation of different domains: poetry (item 1), fiction (item 7), music (item 4), architecture (item 11) and nature (item 15).

Lastly, a set of five items were added to assess behavioral indicators of aesthetic responsiveness; one assessing attendance to museums or performances (item 2) and four probing levels of creative behavior across the domains of writing (item 9), visual arts (item 6), music (item 4) and education (item 12), which we assume to be strongly related to aesthetic responsiveness. To record and score responses, a frequency scale with five categories from "never" to "very often" was implemented. A full list of the 18 items of the original version can be found in the online supplemental material. In sum, aesthetic responsiveness was operationalized as an individual's perceived frequency of aesthetic experiences as indicated by a variety of cognitive and affective states, responses, and behaviors.

This scale construction process emphasizes both, a common origin of aesthetic responses (i.e. aesthetic responsiveness), and multiple facets of aesthetic responsiveness, namely appreciation of aesthetic stimuli, intense aesthetic experiences, and creative behavior. However, it is important to note that the construction of the assessment instrument and its empirical applications were not intended to explore qualitatively different theoretical models of aesthetic experience and its precursors, moderators, mediators, and consequences; or to compare aesthetic responsiveness with aesthetic sensitivity; or to differentiate theoretically refined constructs of the aesthetic process such as aesthetic appreciation, engagement, or taste. The level of detail required for such an investigation and subsequent analysis of the nomological network is beyond the scope of this paper.

With the aim of broadening the applicability of this scale, all items were translated to German language by two bilinguals following widely used guidelines (van de Vijver & Hambleton, 1996). Translations were discussed with one of the developers of the English language original scale with regard to differences and similarities in semantic content. The resulting German language version was used in several research projects at the Max Planck Institute for Empirical Aesthetics in Frankfurt am Main, Germany.

The major aims of this study were (a) to explore and confirm the dimensionality of the scale; (b) to test for measurement invariance of the resulting scale across the English and German language versions; (c) to report scale score descriptive statistics and estimate the reliability of scores of the final scale; and (d) to explore the validity of scale scores using measures of constructs related to aesthetic responsiveness, and investigate associations with responses to specific aesthetic stimuli, namely visual art, poems and music.

Method

Samples

U.S. sample. 285 undergraduate students filled in the scale as part of a battery of tests and questionnaires administered at the beginning of an introductory psychology course at New York University. The battery was completed as an online web survey within the first week of the semester at a time and place of the participants' choosing. Consent was obtained via an online consent form, and all study procedures were approved by the NYU institutional review board. Four cases were excluded as they did not provide any data on the scale. Thus, the final sample comprised 281 participants, 198 (70%) females. The mean age of participants was 18.9 years (*SD* = 1.1), ranging from 16 to 24 years. One missing item response from one participant was imputed using the item sample mean. All participants had completed high-school.

German sample. The German sample consisted of two subsamples. German subsample 1 was a convenience sample of participants from a study on music listening behavior. For this study, 202 participants were recruited, of which 31 did not provide any responses on the aesthetic responsiveness scale, and one had 78% missing responses. Removing these participants resulted in a final sample of 170 participants, 118 females (69%) (7 participants, 4%, did not respond), with a mean age of 31.1 years (*SD* = 12.5; range: 18 to 75 years); 73 (43 %) had completed a university degree.

German subsample 2 was a convenience sample from a study of poem reading. After the reading study, participants filled in the aesthetic responsiveness scale as part of a larger set of questions. The sample consisted of 123 participants, 92 (75%) females, with a mean age of 25.0 years (*SD* = 5.1; range: 18 to 43 years); 54 (44 %) completed a university degree.

German subsamples 1 and 2 were pooled into a German total sample comprising 293 participants, 210 (72%) females (7 participants, 2%, did not identify as one of the sexes), with a mean age of 28.3 years (*SD* = 10.7).

In addition, the final version of the AReA was applied in a validation study comprising 207 participants, 124 (60%) females (1 participant, 0.5% did not identify as one of the sexes), with a mean age of 49.9 years (*SD* = 16.2).

Adding up across countries, the total sample size for this study was *N* = 781.

Measures

All participants filled in the 18 items of the original version of the aesthetic responsiveness scale, except for validation study 4 where the final 14-item version was filled in. In addition, we used responses on sample-specific scales relevant for validation of the AReA. Measures used for validation studies are described in the respective sections.

Data analysis

Item development aimed at emphasizing a common factor underlying responses to all items on the one hand, and multifacetedness of responses with regard to general appreciation, intensity, and creativity, on the other hand. We therefore first analyzed heterogeneity of the items using basic item characteristics such as item-rest correlations (IRC) and inter-item correlations to eliminate single items that clearly did not show satisfactory associations with the other items and were therefore not compatible with the assumption of a single common factor. With the aim of identifying items with invariant measurement characteristics in both samples, this was done separately for the US and the German sample. We then split the

sample randomly by language version into two subsamples, each comprising half of the US and German total sample (random sample 1 and 2; n = 287 each). Using random sample 1, the remaining items were subjected to a parallel analysis based on principal components analysis (PCA) to explore potential dimensional heterogeneity and determine the number of factors to be extracted. We extracted the number of factors estimated ±1 (cf. Lim & Jahng, 2019) and subjected the items to a maximum-likelihood exploratory factor analysis (EFA) with oblique oblimin rotation. We evaluated solutions on the basis of interpretational validity and clarity of the simple structure of rotated factor loadings.

To check for stability of the factorial structure across random samples, we tested second-order confirmatory factor analysis (CFA) models in random sample 2. If the EFA suggested a multiple factor solution, these factors were represented in the CFAs as first-order factors which loaded on a common second-order factor Aesthetic Responsiveness. For testing fit of the factorial structure in random sample 2, we ran the following model sequence: First, we tested CFA models separately in the US and German sample to evaluate if the factorial structure showed an acceptable fit in each language version. We used comparative fit index (CFI) and Tucker-Lewis index (TLI) close to .95 or higher, a standardized root-mean-square residual (SRMR) close to .08 or lower, and a root-mean-square error of approximation (RMSEA) close to .06 or lower, as targets for acceptable model fit in accordance with Hu and Bentler (1999). We then proceeded to test for configual, metric, and scalar measurement invariance (Chen et al., 2005; Millsap, 2011) between the English and German language versions of the scale by comparing model fit for the US sample and the pooled German sample from random sample 2. Configural invariance assumes equal factorial structures in both groups. For model identification, the loading of the first measured variable on each latent factor was fixed to one, the latent common first-order factor means fixed to zero, and intercepts, latent factor variances and covariances freely estimated. Metric invariance

additionally assumes equal factor loadings in both groups. Model specification was the same as for the configural invariance model, except that, first, all first-order factor loadings were constrained to be equal across groups; second, all second-order factor loadings were constrained to be equal. Scalar invariance additionally assumes equal item intercepts. Model specification was the same as for the metric invariance model, except that, first, all item intercepts were constrained to be equal across groups, and the second-order latent factor mean was freely estimated in the German sample, and, second, the second-order factor mean was constrained to be equal between the groups. If one of the invariance assumptions did not hold, we tested for partial invariance by relaxing equality constraints for those parameters that showed substantial modification indices.

Although we report chi-square differences ($\Delta\chi^2$) for all model comparisons, our decisions on measurement invariance were based on differences in approximate fit indices, as $\Delta\chi^2$ is highly sensitive to sample size. In particular, differences in CFI (ΔCFI), RMSEA (ΔRMSEA), and SRMR (ΔSRMR) between models with increasing restrictions were used to assess each level of measurement invariance. In the case of metric invariance, changes of $\Delta\text{CFI} \leq -.010$, $\Delta\text{RMSEA} \geq .015$, and $\Delta\text{SRMR} \geq .015$ would indicate non-invariance as suggested by Cheung and Rensvold (2002) and Chen (2007). In the case of scalar invariance, $\Delta\text{SRMR} \geq .010$ would indicate non-invariance, with the other criteria being the same as for metric invariance, as suggested by Chen (2007).

We then compared factor scores and scale mean scores between language versions in the combined random samples. Note that factor scores, i.e. latent mean differences, can be meaningfully compared between groups even in the case of partial scalar invariance, whereas composite scores (i.e. differences of mean or sum scores) are biased if full measurement invariance does not hold (Steinmetz, 2013). Nevertheless, studies applying psychometric scales often prefer composite scores over factor scores. Composite reliability was separately

estimated for the two versions using coefficient omega (McDonald, 1999), which is appropriate for unit-weighted scoring of congeneric scales (McNeish, 2018). Finally, we investigated construct validity of the resulting scale using Pearson correlation coefficients with relevant experimental data and other self-report scales related to the construct of aesthetic responsiveness.

All models were based on continuous indicator variables using a maximum likelihood estimator with standard errors and a mean-adjusted χ^2 test statistic (MLM) that are robust to non-normality of indicator variable distributions.[1] CFAs and composite reliability calculations were performed using Mplus (Version 7.3); EFAs, parallel analysis, factor extraction and rotation, item, scale and some validity anaylses were performed using Stata (Version 15.1); the remaining validity analyses were performed using *R* (Version 3.4.0).

Results

Item selection and factor analyses

Although the items were designed to indicate different facets of a disposition to respond to aesthetic stimuli, we assumed that they share variance attributable to a common underlying factor, i.e. aesthetic responsiveness. We therefore expected all items to show relatively high associations with the scale score minus the item itself, i.e. IRC, and at least medium inter-item correlations. Sample-specific IRCs as well as average inter-item correlations were higher in the English language version than in the German language version (see Tables S1 and S2 in the supplementary material for details); three items showed very weak IRCs of less than .30 in the German language version, one of which was also very weak in the English language version. We therefore excluded these items (number 7, 15, and 17 of the original scale, cf. Tables S1/S2) from the scale. This increased the average inter-item

[1] We have also tested CFA models for ordered-categorical factor indicators separately for the English and German language version. As these models yielded similar fit to the data as the models for continuous indicators, we used the more straightforward continuous indicator CFA models for measurement invariance analysis.

correlations considerably to .46 in the English language and to .35 in the German language version, bringing the whole scale closer towards a more homogenous item sample.

The resulting 15 items were subjected to a parallel analysis using random sample 1 (both language versions together). Parallel analysis suggested extraction of two factors (Eigenvalues PCA: 6.91; 1.37; 1.09; Eigenvalues parallel analysis: 1.41; 1.32; 1.25). We therefore compared rotated factor solutions with one, two, and three factors. Both, the two- and three-factor solutions clearly separated a creative behavior factor. The three-factor solution provided a clearer simple structure and an interpretable third factor, although one item did not fit with the content of the creative behavior factor despite a high factor loading. This was likely due to confounding content ("I enjoy poetry", while poetry and writing was also prominently represented in two other items loadings on the creative behavior factor). We therefore decided to remove this item and rerun the analysis, resulting in a clear and interpretable simple structure with three factors. Factor 1 represented aesthetic appreciation, factor 2 strong/intense emotional responses to art exposure, and factor 3 different aspects of producing art. One item ("I am deeply moved when I see art") cross-loaded on the factors representing aesthetic appreciation and intense aesthetic experience. The correlations between the factors were: $r_{f1,f2} = .67$, $r_{f1,f3} = .48$, $r_{f2,f3} = .46$.

To check stability of the factorial structure across random samples, we conducted second-order CFAs using random sample 2. CFA models were fitted separately for the English and German language versions. The CFA model showed an acceptable fit to the data in both, the English language ($\chi^2 = 112.6$; df = 73; p = .002; RMSEA = 0.062, 90% CI: 0.038, 0.084; CFI = 0.965; TLI = 0.957; SRMR = 0.050) and German language version ($\chi^2 = 119.6$; df = 74 (the residual variance of one first-order factor in the German sample had a small negative estimate and was therefore set to zero); p = .001; RMSEA = 0.065, 90% CI: 0.042, 0.086; CFI = 0.946; TLI = 0.933; SRMR = 0.050). These results provide support for the

validity of the factorial structure across different samples.

In sum, the 3-factor model provided the best mixture of good model fit, parsimoniousness, and interpretability, and it was confirmed in an independent random sample using second-order CFAs. The final scale was named *Aesthetic Responsiveness Assessment* (AReA), comprising the sub-scales Aesthetic Appreciation (AA), Intense Aesthetic Experience (IAE), and Creative Behavior (CB), loading on a second-order factor Aesthetic Responsiveness (AReA total). Both language versions of the final scale can be found in the supplementary material to this article.

Measurement invariance across language versions

We tested the final second-order CFA model for configural, metric, and scalar measurement invariance across the English and German language versions using the US and the pooled German sample. As can be seen from Table 1, the configural invariance model yielded acceptable model fit indices. Comparing fit indices of the model with equal first-order factor loadings to the configural invariance model showed that changes of RMSEA, CFI, and SRMR were minimal and within or close to the pre-defined cut-off values. In addition, all model fit indices suggested a good fit of the metric model. The second-order metric invariance model showed very small deviations from the first-order meric invariance model. We therefore concluded that these results clearly suggest full metric invariance across the English and German language versions of the AReA. In contrast, the test of scalar invariance of observed indicators yielded model fit indices that were clearly beyond pre-defined cut-off values for model fit as well as fit difference to the metric invariance model. Inspection of modification indices suggested that this was due to item intercept equality constraints for few items. Lifting equality constraints for three items (see Table 1 for details) resulted in an acceptable model fit as well as fit-index differences that were within or very close to the pre-defined range for demonstrating scalar invariance of observed indicators. Testing scalar

invariance of first-order factors showed very small deviations from the observed-indicator scalar invariance model. These results suggest that the English and German language versions of the AReA showed partial scalar invariance.

Figure 1 shows structure and coefficients of the final partial scalar measurement invariance model. The good fit of the second-order CFA model supports the assumption of a single higher order factor explaining the covariance between the first-order factors. We therefore suggest that scoring of the AReA should, in addition to computation of scores for the three factors, also include computation of a total score reflecting individual aesthetic responsiveness.

Fitting the CFA model shown in Figure 1 to data from another German validation sample of 207 participants resulted in a good model fit ($\chi^2 = 110.1$; df = 73; p = .003; RMSEA = 0.050, 90% CI: 0.029, 0.068; CFI = 0.958; TLI = 0.948; SRMR = 0.052). Factor loadings and latent factor correlations (not shown here) were similar to the results for random sample 2 shown in Figure 1. These results further support the factorial validity of the AReA German language version.

Scale scores

Table 2 shows average scale mean scores for the US and the German total samples. Although some of the scale score distribution tests indicated slight deviations from normality, the absolute skewness and kurtosis parameters as well as inspection of histograms showed that these deviations were minor. As factor scores from the partial scalar measurement invariance model can be used for unbiased comparison of individual trait standings between language versions, we computed correlations between factor scores and scale mean scores. These correlations were very high (Table 2), supporting the utility of scoring the AReA using sum or mean scale scores.

Reliability

Composite reliability coefficients were all in a satisfactory range of ω > .70 for both language versions (cf. Table 2). Coefficients were slightly higher in the US sample, with the exception of the subscale CB. Notably, CB yielded acceptable reliability estimations despite comprising only three items.

Results of reliability analysis in the additional German validation sample of 207 participants suggested good reliabilities for the AReA total scale (ω = .82) and the subscales AA (ω = .84) and IAE (ω = .80). In contrast, the reliability estimate for the subscale CB was somewhat lower (ω = .63), both in comparison with the other AReA subscales in this sample, and in comparison to other samples (cf. Table 2).

Validation study 1: Trait pleasure and responses to visual artworks and music

The US validation sample consisted of an independent sample of *n* = 50 participants (mean age = 27.3 yrs., *SD* = 6.5; 19 males, 31 females) who participated in either a study with visual artworks (Belfi et al., 2019) or with musical excerpts. In addition to the AReA, all participants completed the Temporal Experience of Pleasure Scale (TEPS; Gard et al., 2006). The TEPS consists of two sub-scales: TEPS-A, which measures anticipatory pleasure (related to reward-sensitivity and imagery), and TEPS-C, which measures consummatory pleasure (related to openness to diverse experiences and appreciation of positive stimuli). Moreover, aesthetic judgement ratings were available for visual artworks (*n* = 21) and musical excerpts (*n* = 26).[2]

For the TEPS, we expected both scales to show a positive relationship to the AReA sub-scales AA and IAE. Specifically, the TEPS-C scale should bear a positive relationship with the AReA sub-scales, because openness to experience is conceptually closely linked with aesthetic responsiveness. The results shown in Table 3 largely match these expectations,

[2] Note that these two subsamples do not add-up to *n* = 50, because data of three participants had to be discarded due to problems with performance and recording of the aesthetic judgements.

although the TEPS Anticipatory Pleasure scale was only very weakly related to IAE and the AReA total score.

For the visual study, a squeeze ball was used to record continuous momentary aesthetic pleasantness of visual artworks presented for either 1 second, 5 seconds, or 15 seconds. Artworks consisted of 30 paintings at each duration (90 total), selected to represent a variety of styles, content and periods (15th century to present day, Western and Eastern, representational and abstract). Observers were instructed to squeeze the ball at a level corresponding to their felt pleasure both during the painting presentation and for a "post-stimulus" period after the painting disappeared. In addition, participants provided a retrospective overall rating of how aesthetically appealing each trial was using a trackball in the other hand.

For the magnitude of the momentary online and retrospective ratings of visual aesthetic stimuli we expected positive correlations with the AReA sub-scales, again particularly AA and IAE. In this context, associations with online-ratings (i.e., the average and maximum ratings via the squeeze ball during the exposure to the stimuli) should prove more reliable compared to associations with retrospective ratings, as they better reflect the momentary experience, whereas retrospective measures are potentially biased. In addition, the maximum rating might show stronger relations to the AReA sub-scales, because they provide an index of the maximum reactivity of a participant. As we expected that exposure to an artwork for the duration of merely one second is substantially too short to provoke a reliable aesthetic response, we compared associations of AReA subscales with ratings during 1-second exposure separately from ratings during 5- and 15-second exposure.

For the sample of participants that received visual stimuli, Table 4 provides correlations between the average and maximum online-ratings, and the retrospective ratings for 1 second duration exposure and 5 and 15 second duration exposure with the AReA sub-

scales. As can be seen, AReA values were not predictive of aesthetic judgments in the 1-second exposure conditions, but correlated with aesthetic judgments in the longer conditions. However, this was only the case for momentary online ratings, but not for retrospective ratings. Moreover, there was a tendency for stronger relations to the maximum online ratings compared to the average online ratings.

For the auditory study, participants listened to 60 s excerpts of music and made continuous ratings of liking on a 0 (Low) to 1 (High) visual slider scale using a trackball. Following each clip, observers gave an overall rating of how aesthetically appealing the clip was. Clips consisted of 16 classical pieces and 16 electronic pieces, blocked by genre in groups of 8 clips. Within these genres, pieces were selected to be stylistically consistent in order to prevent participants from responding purely on the basis of genre. Classical pieces were of 19th century small ensemble music from the Romantic era, which contains a wider range of dynamic and emotional intensity than other periods. Electronic music consisted of dance music with a distinctive beat structure (60-150 bpm), selected to have some degree of change or transition during the clip; songs with a single repetitive motif were avoided.

For the sample of participants that received music stimuli, Table 5 provides correlations between the average and maximum online-ratings, and the retrospective ratings for classical or electronic music with the AReA sub-scales. As can be seen, AReA scores were substantially correlated with rating of classical music, even though these correlations were not statistically significant due to the small sample.

Validation study 2: Responses to poems

The second German validation sample consisted of a sub-set of n = 40 participants of the German subsample 2, where the effects of rhetorical language features on the subjective aesthetic experience of the reader was investigated (Menninghaus, Wagner, Wassiliwizky, et al., 2017). Participants read 10 poems in their original version and 10 poems in a de-

rhetorized version. Additionally, all participants filled in the AReA and provided ratings of different versions of poems on a 7-point scale for beauty, movingness, melodiousness, joy, and sadness. Previous research on poem and proverb reading has shown that manipulations of rhyme and meter lead to changes in the processing and aesthetic evaluation of language (Menninghaus, Bohrn, et al., 2015; Menninghaus & Wallot, 2020; Wallot & Menninghaus, 2018).

Because AReA is an instrument designed to assess a person's responsiveness to aesthetic stimuli, we hypothesized that participants scoring high on the AReA would provide higher ratings on subjective emotional and aesthetic experience for the original poems compared to participants that scored low on AReA. Additionally, we hypothesized that participants scoring high on the AReA would show a greater difference between original poems and their de-rhetorized versions (i.e., without rhyme and meter), indicating greater sensitivity to the absence vs. presence of those poetic language features. The subscales Aesthetic Appreciation and Intense Aesthetic Experience were expected to show stronger associations in contrast to Creative Behavior.

Table 7 shows the correlations between the three AReA subscales and the AReA total score with ratings of joy, sadness, beauty, movingness and melodiousness. The average ratings correlated consistently positively with the Intense Aesthetic Experience subscale, and less so with the Creative Behavior subscale. However, in contrast to our hypothesis, only values for beauty ratings correlated positively with the Aesthetic Appreciation subscale. For the difference scores, we found significant positive correlations on three out of the five ratings for the Intense Aesthetic Experience subscale, but none for the other two subscales. While these results support the validity of the AReA, it seems that responses to poetry are more strongly affected by a disposition to intense aesthetic experiences as assessed by the IAE subscale of the AReA.

Validation study 3: Behavioral activation, music reward, and responses to music

The first German validation sample consisted of the whole sample of $n = 167$ participants of the German subsample 1, drawn from a study on evaluating listeners' responses to music in order to identify individuals who show low levels of hedonic pleasure during music listening. In addition to the AReA, participants filled in the German version of the BIS/BAS (Carver & White, 1994; Strobel et al., 2001), and a German ad-hoc translation of the Barcelona Music Reward Questionnaire (BMRQ; Mas-Herrero et al., 2013), and were asked to rate how often they experience chills during music listening in general (possible answers: 1 = "never", 2 = "rarely", 3 = "sometimes", 4 = "often"). In addition, participants were asked to listen to a piece of music that had been selected for reliably eliciting chills across a majority of listeners. Afterwards, participants were asked to rate whether they experienced chills while listening to the given piece of music (possible answers: 1 = "no", 2 = "yes", or 3 = "don't know"). For the latter variable, we removed "don't know" answers before analysis.

The BIS/BAS consists of the following sub-scales: The BIS total score (sensitive to signals of punishment, non-reward and novelty), the BAS total score (sensitive to signals of reward, non-punishment and escape from punishment), as well as three BAS-subscales: BAS-Drive (pursuit of desired goals), BAS-Fun-Seeking (desire for new rewards and willingness to approach), and BAS-Reward (positive responses to occurrence or anticipation of reward). Because AReA was designed to assess a person's sensitivity to aesthetic stimuli primarily relating to a (positive) emotional response, we hypothesized the following: In relation to the AReA subscales, there should be no particular relation to the BIS total score, as AReA items are not related to negative experiences or their avoidance. In contrast, we expected positive associations with the BAS total score, and particularly with the BAS-Reward subscale, as aesthetic experiences are rewarding. As the BIS/BAS captures strong emotional responses, we

expected strong positive associations with the AReA subscale Intense Aesthetic Experience, but to a lesser degree to Aesthetic Appreciation.

The BMRQ consists of five subscales: BMRQ-Musical-Seeking (e.g. looking out for new music, informing oneself, spending money), BMRQ-Emotional-Evocation (e.g chills, tears, becoming emotional), BMRQ-Mood-Regulation (e.g. keeps me company, helps me relax), BMRQ-Sensory-Motor (e.g. need to dance, tap, sing, hum), BMRQ-Social-Reward (e.g. like to play with others, feeling of connection). In relation to the AReA subscales, we expected positive associations with the BMRQ-Emotion-Evocation subscale, which should tap into the same construct as the AReA Aesthetic Appreciation and Intense Aesthetic Experience subscales. Furthermore, the subscale BMRQ-Sensory-Motor seems to be unrelated to the AReA subscales, because it neither captures any form of evaluation of emotional involvement, nor a productive component in the sense of the Creative Behavior subscale. Associations between the other three subscales of the BMRQ and AReA were difficult to predict, because even though they do emphasize emotional components of music perception, they additionally capture consequences of functions of listening to music that are not specifically addressed in the AReA. Finally, the two chill variables were expected to be positively associated with the AReA subscales Aesthetic Appreciation and particularly Intense Aesthetic Experience, because chills are a bodily response indicative of high physiological arousal (Wassiliwizky et al., 2017) triggered by stimuli with high information content (Omigie et al., 2019)

Table 6 shows the correlations between the three AReA scale scores and the subscales of the BIS/BAS, the BMRQ, and ratings of occurrence of chills (trait and state). The hypothesized relations are generally borne out: Specifically, the AReA subscales did not correlate with the BIS total score of the BIS/BAS and the Sensory-Motor score of the BMRQ. Furthermore, the Creative Behavior subscale of the AReA showed the smallest correlations

with all other measures that were expected to be more strongly associated with the receptive subscales of the AReA. Particularly, the hypothesized positive correlations between the AReA subscales Aesthetic Appreciation and Intense Aesthetic Experience with the BAS Reward subscale, BMRQ Emotional Evocation subscale, and trait and state measures of chills were observed.

Validation study 4: Big Five, open-mindedness and its facets

In another German validation sample, an online survey presented the final 14-item AReA version as well as a German translation of the BFI-2 (Danner et al., 2019; Soto & John, 2017) and was completed by 207 participants (3 participants were excluded due to extremely long response times). We computed Pearson's correlation coefficients between AReA scale scores and the BFI-2 domain scales as well as the three facet scales constituting Open-Mindedness, i.e. Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination. The pattern of correlations will provide additional information on the convergent and discriminant validity of the AReA scales. We expected large correlations between AReA scales and the Open-Mindedness scale, but much smaller correlations with the other domain scales, i.e. Extraversion, Agreeableness, Conscientiousness, and Negative Emotionality. With regard to the facet scales of Open-Mindedness, large correlations with AReA scales were expected for the facet Aesthetic Sensitivity, whereas correlations with the other facet scales were expected to be much smaller. Finally, the correlation between the AReA subscale Creative Behavior and the facet scale Creative Imagination was expected to be higher than with the other facet scales, as an individual disposition to high levels of creative imagination is expected to facilitate creative behavior as assessed by the AReA subscale.

Table 8 shows correlations between AReA and BFI-2 scales. As expected, correlations of the AReA with Open-Mindedness were large and highly significant, whereas those with Agreeableness, Conscientiousness, and Negative Emotionality were small and mostly not

significantly different from zero. Extraversion showed significant positive correlations with the AReA scales, due to a considerable portion of shared variance between Extraverion and Open-Mindedness (r = .36). However, these correlations were significantly smaller than the correlations between AReA scales and Open-Mindedness (difference tests for correlation coefficients: all ps ≤ .001, see supplemental Table S3 for details). Regarding the facets of Open-Mindedness, the AReA subscales correlated significantly higher with the facet Aesthetic Sensitivity than with the other facets (ps < .05, see supplemental Table S3 for details), with the exception of the AReA subscale Creative Behavior. In line with our expectations, CB showed significantly higher correlations with Creative Imagination than with the other facets (all ps ≤ .020, see supplemental Table S3 for details).

In summary, results of validation study 4 support factorial, convergent, and discriminant validity of the AReA total and subscale scores in its German version, and therefore further strengthen the evidence for construct validity of the AReA.

Discussion

We present the Aesthetic Responsiveness Assessment (AReA) which can be used to assess aesthetic responsiveness. The scale is based on an original pool of questionnaire items that was compiled with the goal of identifying potential study participants that are particularly responsive to aesthetic stimuli. The final version comprises three sub-scales: Aesthetic Appreciation (AA), Intense Aesthetic Experience (IAE), and Creative Behavior (CB) of respondents.

A main goal of the scale was to allow experimenters to distinguish those individuals who regularly respond to artworks in an intense way from those who rarely experience more than a commonplace appreciation of aesthetic objects in everyday life. In supporting the notion that such a distinction is an important one to make, our scale complements previous scales, such as the EBS (Diessner et al., 2008), which focused on other distinctions (e.g.

between reponses to nature, art and moral beauty).

Indeed, the dissociation of the two reception-oriented sub-scales AA and IAE fits with previous behavioral findings on the special capacity of engagement with art to result in intense aesthetic experiences such as being moved (Menninghaus, Wagner, et al., 2015). This dissociation is in line with neurophysiological findings showing that prefrontal and default mode network brain regions are selectively engaged by strongly moving aesthetic experiences with visual artwork (Belfi et al., 2019; Vessel et al., 2012, 2013). Similarly, it is in line with evidence that experiences of beauty in reponse to music may vary in terms of subjective and physiological arousal (Omigie et al., 2019). The extraction of the CB subscale clearly reflects item content relating to participants' engagement in the creation of art. We suggest that this makes it highly relevant for occasions when it is important to identify participants that regularly engage in the production of art works. However, in contrast to high reliabilities of the AReA total scale score and scores on AA and IAE, the shortness of the CB scale limits its reliability, which implies a relatively larger measurement error in the assessment of individuals. This should be kept in mind when using the CB scale as a screening tool for selection of individuals.

One of the most important findings is the demonstration of measurement invariance for the English and German language versions of AReA. Having established full metric invariance suggests that results of association analyses such as regression using the AReA scales can be meaningfully compared between samples from Germany and the US using the respective language versions. However, one should be cautious when comparing mean levels of responses (i.e. composite scores) across English and German language versions, because full scalar invariance had to be rejected for this instrument. Thus observed differences between the samples cannot be fully attributed to differences in individual latent trait standing. However, partial scalar invariance was found when item intercept equality

constraints were released for three items from the scales AA and CB. Hence, analyses of composite differences between language versions of the AReA or its subscales AA and CB should use factor scores, i.e. latent mean differences (Steinmetz, 2013), while composite scores can be compared between language versions when analyzing IAE subscale scores only.

Using independent samples or sub-samples of participants that took part in different studies on the reception and evaluation of music, visual art, and poetry, we found evidence supporting the validity of scale scores by showing expected correlations with self reported strength of aesthetic responsiveness to visual (validation study 1), musical (validation studies 1 and 2) and literary aesthetic stimuli (validation study 3), as well as scales tapping into general (BIS/BAS and TEPS), and more domain-specific hedonic responses (BMRQ). Although due to small sample sizes not all of these correlations were statistically significant, many of them represent rather large effects from a normative perspective (Gignac & Szodorai, 2016). These results suggest a broad applicability of AReA as a screening instrument across a variety of domains of art perception.

As there is considerable overlap between the construct of aesthetic responsiveness and the personality domain opennenss, relatively high correlations between measures of these constructs should be expected. The pattern of correlations of the AReA with measures of the Big Five personality domains and the facets of Open-Mindedness we found in validation study 4 were in line with these expectations. The large correlations between the Open-Mindedness facet Aesthetic Sensitivity and AReA scales support its convergent validity. However, the size of the correlations clearly suggests that the constructs measured by the AReA are suffiently different to support its utility as an independent measurement instrument. This is further supported by the specific association of CB with Creative Imagination. In contrast, AReA scale scores did not correlate substantially with agreeableness, conscientiousness, and negative emotionality, while the moderate correlations with

extraversion are likely due to shared variance with opennenss. In total, these results strongly support the construct validity of the AReA in its German language version, and they can be expected to generalize to the English language version, as the measurements are invariant across languages. Nevertheless, future studies should investigate similar correlations using an English speaking sample.

We conclude that AReA scores indicate the theoretical construct of aesthetic responsiveness. Our theoretical approach emphasizes the individual subjective experience associated with central processing of aesthetic stimuli. Similar to what has been suggested in the area of stress reactivity (Schlotz, 2013; Schlotz et al., 2011), it implies relatively consistent and coherent responses across time, stimulus domains, and response domains. As this is a rather strong assumption, future studies should systematically assess and compare responses across domains to put these theoretical assumptions to the test. The development of an inventory that systematically assesses responses in different domains would be a valuable contribution.

It is not surprising that scores on the AReA subscale Creative Behavior (CB) correlated less often and less strongly with judgments of beauty, pleasantness, or aesthetic appeal in reception-oriented tasks than the other two scales, as creative behavior includes an action-related component beyond simply responding to aesthetic stimuli. It could thus be debated whether CB is part of the construct of aesthetic responsiveness in a strict sense. However, we opted to keep this subscale in the AReA, as it provides useful information at relatively low cost (three items only) on an important aspect of aesthetics; namely a predisposition to engage in art production. Indeed, both, substantial correlations between factors, and good fit of the second-order CFA model provide psychometric evidence that supports keeping CB as a subscale of the AReA.

It should be noted that theoretically, aesthetic responsiveness includes both indicators

of aesthetic appreciation and aesthetic engagement. Both are assumed to be affected by an individual's trait standing on aesthetic responsiveness. Consequently, the AReA does not separate these constructs systematically (although the subscale Aesthetic Appreciation contains less engagement-relevant items than the other subscales). The relative contribution of aesthetic responsiveness to appreciation and engagement could differ between individuals (individual-specific response patterns), and probably even within individuals across time or stimuli. However, a theoretical conception that separates individual propensities to aesthetic engagement vs. appreciation—as two related but separable facets of aesthetic responsiveness—is not incompatible with our theoretical account of aesthetic responsiveness. Future developments of assessments of aesthetic responsiveness could aim at generating items that more systematically sample specific theoretically defined components of aesthetic responsiveness. One approach could be a systematic separation of aesthetic appreciation and aesthetic engagement. Another one could be a differentiation of response indicators to more specifically reflect emotional, cognitive, behavioral, and physiological domains. Whether such refinements of the operationalization of aesthetic responsiveness have utility and incremental validity compared to the AReA is an empirical question.

It is important to note that the construct of aesthetic responsiveness explicitly excludes reference to an external standard and is therefore very different from constructs that refer to quality of judgements of aesthetic stimuli such as aesthetic sensitivity (Child, 1964; Eysenck, 1940; Myszkowski & Zenasni, 2016; but see Corradi et al., 2019). This has the great advantage that the AReA can be used in non-experts and experts alike. Our theoretical approach clearly implies that the question of whether these groups differ in their aesthetic responsiveness is not a theoretical but an empirical issue. However, the construct defined here nevertheless refers to responsiveness to aesthetic stimuli, and any measure of the construct has to demonstrate that scores reflect more than just non-specific responsivity. In this sense,

our finding from validation study 2 that AReA scores correlated more strongly with responses to classical versus electronic music can be seen as a first step towards specificity of responsiveness to aesthetically relevant stimuli.

Limitations and outlook

There might be certain limitations built into the convenience samples that were used in the current analysis. For example, some studies have found differences in art perception and consumption between experts and laypersons (Elvers et al., 2015; Leder et al., 2014). As our samples comprised laypersons, its properties in a sample of experts might be different. To clarify this point, a future study could investigate measurement invariance of the AReA between laypersons and experts.

Also, there is a certain built-in limitation of the scale with regard to the original item pool of the screening instrument: Currently, the items of the scale focus disproportionally on wordings that are suggestive of visual perception of art, especially compared to other domains such as music and literature (or nature). Even though the results of our validation studies suggest that the scale can successfully be applied to those domains, it does not provide a fine-grained distinction between domains. Moreover, the current item pool does not systematically cover response domains. For example, IAE captures emotional and physiological responses, but it does not distinguish between them, and does not comprise items indicating other response domains. Hence, future developments should include a more systematic selection of additional items from different aesthetic and response domains to provide a more fine-grained instrument, potentially also covering negative emotional responses to art (Menninghaus, Wagner, Hanich, et al., 2017). Finally, it might be of interest to explore what background experiences lead to high scores on the AReA. More specifically, it would be interesting to investigate the relative contribution of frequency and intensity of individual aesthetic experiences to scores on the AReA.

The mixture of exploratory and confirmatory strategies in the construction of the AReA resulted in a stable and meaningful scale structure. However, alternative structures are conceivable that emphasize other aspects of aesthetic responsiveness theory. Such alternative operationalizations could be based on refined theoretical accounts and would provide potentially useful progress in the assessment of aesthetic responsiveness. In addition, multimodal assessments of responses could provide insight into aesthetic responsiveness beyond self-reports.

Conclusion

Although built on an exploratory scale construction strategy, the AReA is a promising, psychometrically evaluated tool for the assessment of individual differences in aesthetic responsiveness that is particularly suitable for selecting participants for empirical aesthetics studies. It can also be used to study (a) associations of aesthetic responsiveness with other constructs from the area of aesthetic research such as aesthetic sensitivity, (b) associations with constructs from the broader area of personality, such as personality dimensions or ability, and (c) developmental trajectories and factors underlying individual aesthetic responsiveness. As we demonstrated measurement invariance for the AReA, its English and German language versions can be used in parallel to compare samples between these languages.

References

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150-166. https://doi.org/10.1177/1088868306294907

Belfi, A. M., Vessel, E. A., Brielmann, A., Isik, A. I., Chatterjee, A., Leder, H., Pelli, D. G., & Starr, G. G. (2019). Dynamics of aesthetic experience are reflected in the default-mode network. *Neuroimage, 188*, 584-597. https://doi.org/10.1016/j.neuroimage.2018.12.017

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. Appleton-Century-Crofts.

Brielmann, A. A., & Pelli, D. G. (2017). Beauty requires thought. *Current Biology, 27*(10), 1506-1513.e3. https://doi.org/10.1016/j.cub.2017.04.018

Burke, E. (2015). *A philosophical enquiry into the sublime and beautiful*. Oxford University Press. (Original work published 1757)

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319-333. https://doi.org/10.1037/0022-3514.67.2.319

Chatterjee, A., & Vartanian, O. (2016). Neuroscience of aesthetics. *Annals of the New York Academy of Sciences, 1369*(1), 172-194. https://doi.org/10.1111/nyas.13035

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal, 12*(3), 471-492. https://doi.org/10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. https://doi.org/10.1207/s15328007sem0902_5

Child, I. L. (1964). Observations on the meaning of some measures of esthetic sensitivity. *Journal of Psychology, 57*(1), 49-64. https://doi.org/10.1080/00223980.1964.9916671

Corradi, G., Chuquichambi, E. G., Barrada, J. R., Clemente, A., & Nadal, M. (2019). A new conception of visual aesthetic sensitivity. Advance online publication. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12427

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*(1), 21-50. https://doi.org/10.1207/s15327752jpa6401_2

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Das Big Five Inventar 2 [The Big Five Inventory 2]. *Diagnostica, 65*(3), 121-132. https://doi.org/10.1026/0012-1924/a000218

Diessner, R., Pohling, R., Stacy, S., & Güsewell, A. (2018). Trait appreciation of beauty: A story of love, transcendence, and inquiry. *Review of General Psychology, 22*(4), 377-397. https://doi.org/10.1037/gpr0000166

Diessner, R., Solom, R. C., Frost, N. K., Parsons, L., & Davidson, J. (2008). Engagement with beauty: appreciating natural, artistic, and moral beauty. *Journal of Psychology, 142*(3), 303-329. https://doi.org/10.3200/jrlp.142.3.303-332

Elvers, P., Omigie, D., Fuhrmann, W., & Fischinger, T. (2015). Exploring the musical taste of expert listeners: musicology students reveal tendency toward omnivorous taste. *Frontiers in Psychology, 6*, Article 1252. https://doi.org/10.3389/fpsyg.2015.01252

Eysenck, H. J. (1940). The general factor in aesthetic judgement. *British Journal of Psychology, 31*(1), 94-102. https://doi.org/10.1111/j.2044-8295.1940.tb00977.x

Fayn, K., MacCann, C., Tiliopoulos, N., & Silvia, P. J. (2015). Aesthetic emotions and aesthetic people: Openness predicts sensitivity to novelty in the experiences of interest and pleasure. *Frontiers in Psychology, 6*, Article 1877. https://doi.org/10.3389/fpsyg.2015.01877

Gard, D. E., Gard, M. G., Kring, A. M., & John, O. P. (2006). Anticipatory and consummatory components of the experience of pleasure: A scale development study. *Journal of Research in Personality, 40*(6), 1086-1102. https://doi.org/10.1016/j.jrp.2005.11.001

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74-78. https://doi.org/10.1016/j.paid.2016.06.069

Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new visual aesthetic sensitivity test I: Construction and psychometric properties. *Perceptual and Motor Skills, 49*(3), 795-802. https://doi.org/10.2466/pms.1979.49.3.795

Güsewell, A., & Ruch, W. (2012). Are there multiple channels through which we connect with beauty and excellence? *Journal of Positive Psychology, 7*(6), 516-529. https://doi.org/10.1080/17439760.2012.726636

Hager, M., Hagemann, D., Danner, D., & Schankin, A. (2012). Assessing aesthetic appreciation of visual artworks—The construction of the Art Reception Survey (ARS). *Psychology of Aesthetics, Creativity, and the Arts, 6*(4), 320-333. https://doi.org/10.1037/a0028776

Haidt, J., & Keltner, D. (2004). Appreciation of beauty and excellence [awe, wonder, elevation]. In C. K. Peterson & M. E. P. Seligman (Eds.), *Character strengths and virtues* (pp. 537-551). Oxford Universty Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure

analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55.

Istok, E., Brattico, E., Jacobsen, T., Krohn, K., Muller, M., & Tervaniemi, M. (2009). Aesthetic responses to music: A questionnaire study. *Musicae Scientiae, 13*(2), 183-206. https://doi.org/10.1177/102986490901300201

Jacobsen, T., & Beudt, S. (2017). Domain generality and domain specificity in aesthetic appreciation. *New Ideas in Psychology, 47*, 97-102. https://doi.org/10.1016/j.newideapsych.2017.03.008

Jacobsen, T., Buchta, K., Köhler, M., & Schröger, E. (2004). The primacy of beauty in judging the aesthetics of objects. *Psychological Reports, 94*(3_suppl), 1253-1260. https://doi.org/10.2466/pr0.94.3c.1253-1260

Karwowski, M., & Beghetto, R. A. (2019). Creative behavior as agentic action. *Psychology of Aesthetics, Creativity, and the Arts, 13*(4), 402-415. https://doi.org/10.1037/aca0000190

Lane, R. D., Waldstein, S. R., Chesney, M. A., Jennings, J. R., Lovallo, W. R., Kozel, P. J., Rose, R. M., Drossman, D. A., Schneiderman, N., Thayer, J. F., & Cameron, O. G. (2009). The rebirth of neuroscience in psychosomatic medicine, part I: Historical context, methods, and relevant basic science. *Psychosomatic Medicine, 71*(2), 117-134. https://doi.org/10.1097/PSY.0b013e31819783be

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

Levinson, J. (2012). Musical beauty. *Teorema, 31*(3), 127-135.

Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods, 24*(4), 452-467.

https://doi.org/10.1037/met0000230

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual differences in music reward experiences. *Music Perception, 31*(2), 118-138. https://doi.org/10.1525/mp.2013.31.2.118

McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience. *Motivation and Emotion, 31*(1), 5-11. https://doi.org/10.1007/s11031-007-9053-1

McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.

McManus, I. C., & Furnham, A. (2006). Aesthetic activities and aesthetic attitudes: Influences of education, background and personality on interest and involvement in the arts. *British Journal of Psychology, 97*, 555-587. https://doi.org/10.1348/000712606x101088

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412-433. https://doi.org/10.1037/met0000144

Menninghaus, W., Bohrn, I. C., Knoop, C. A., Kotz, S. A., Schlotz, W., & Jacobs, A. M. (2015). Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition, 143*, 48-60. https://doi.org/10.1016/j.cognition.2015.05.026

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T., & Koelsch, S. (2017). The Distancing-Embracing Model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences, 40*, Article e347. https://doi.org/10.1017/S0140525X17000309

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T. (2015). Towards a psychological construct of being moved. *PLOS ONE, 10*(6), Article e0128451. https://doi.org/10.1371/journal.pone.0128451

Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., & Knoop, C. A. (2017). The

emotional and aesthetic powers of parallelistic diction. *Poetics, 63*, 47-59. https://doi.org/10.1016/j.poetic.2016.12.001

Menninghaus, W., Wagner, V., Wassiliwizky, E., Schindler, I., Hanich, J., Jacobsen, T., & Koelsch, S. (2019). What are aesthetic emotions? *Psychological Review, 126*(2), 171-195. https://doi.org/10.1037/rev0000135

Menninghaus, W., & Wallot, S. (2020). *What the eyes reveal about (reading) poetry* [Manuscript submitted for publication]. Daprtment of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Psychology Press.

Myszkowski, N., Storme, M., Zenasni, F., & Lubart, T. (2014). Is visual aesthetic sensitivity independent from intelligence, personality and creativity? *Personality and Individual Differences, 59*(0), 16-20. https://doi.org/10.1016/j.paid.2013.10.021

Myszkowski, N., & Zenasni, F. (2016). Individual Differences in Aesthetic Ability: The Case for an Aesthetic Quotient. *Frontiers in Psychology, 7*, Article 750. https://doi.org/10.3389/fpsyg.2016.00750

Omigie, D., Frieler, K., Bär, C., Muralikrishnan, R., Wald-Fuhrmann, M., & Fischinger, T. (2019). Experiencing musical beauty: Emotional subtypes and their physiological and musico-acoustic correlates. *Psychology of Aesthetics, Creativity, and the Arts*, Advance online publication. https://doi.org/10.1037/aca0000271

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews, 21*, 80-125. https://doi.org/10.1016/j.plrev.2017.02.003

Peterson, C. K., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook*

and classification. American Psychological Association.

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 59-68. https://doi.org/10.1037/aca0000079

Rawlings, D., Barrantes i Vidal, N., & Furnham, A. (2000). Personality and aesthetic preference in Spain and England: two studies relating sensation seeking and openness to experience to liking for paintings and music. *European Journal of Personality, 14*(6), 553-576. https://doi.org/10.1002/1099-0984(200011/12)14:6<553::AID-PER384>3.0.CO;2-H

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Personality and Socical Psychology Review, 8*(4), 364-382. https://doi.org/10.1207/s15327957pspr0804_3

Rowold, J. (2008). Instrument development for esthetic perception assessment. *Journal of Media Psychology, 20*(1), 35-40. https://doi.org/10.1027/1864-1105.20.1.35

Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., & Scherer, K. R. (2017). Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLOS ONE, 12*(6), Article e0178899. https://doi.org/10.1371/journal.pone.0178899

Schlotz, W. (2013). Stress reactivity. In M. D. Gellman & R. J. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (pp. 1891-1894). Springer.

Schlotz, W., Yim, I. S., Zoccola, P. M., Jansen, L., & Schulz, P. (2011). The perceived stress reactivity scale: Measurement invariance, stability, and validity in three countries. *Psychological Assessment, 23*(1), 80-94. https://doi.org/10.1037/a0021148

Silvia, P. J., Fayn, K., Nusbaum, E. C., & Beaty, R. E. (2015). Openness to experience and awe in response to nature and music: Personality and profound aesthetic experiences.

Psychology of Aesthetics, Creativity, and the Arts, 9(4), 376-384. https://doi.org/10.1037/aca0000028

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117-143. https://doi.org/10.1037/pspp0000096

Stamatopoulou, D. (2004). Integrating the philosophy and psychology of aesthetic experience: Development of the aesthetic experience scale. *Psychological Reports, 95*(2), 673-695. https://doi.org/10.2466/PR0.95.6.673-695

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology, 9*(1), 1-12. https://doi.org/10.1027/1614-2241/a000049

Strobel, A., Beauducel, A., Debener, S., & Brocke, B. (2001). Eine deutschsprachige Version des BIS/BAS-Fragebogens von Carver und White [A German language version of the BIS/BAS questionnaire by Carver and White]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 22*(3), 216-227. https://doi.org/10.1024//0170-1789.22.3.216

van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*(2), 89-99. https://doi.org/10.1027/1016-9040.1.2.89

Vessel, E. A., Isik, A. I., Belfi, A. M., Stahl, J. L., & Starr, G. G. (2019). The default-mode network represents aesthetic appeal that generalizes across visual domains. *Proceedings of the National Academy of Sciences, 116*(38), 19155-19164. https://doi.org/10.1073/pnas.1902650116

Vessel, E. A., Maurer, N., Denker, A. H., & Starr, G. G. (2018). Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition, 179*, 121-131.

https://doi.org/10.1016/j.cognition.2018.06.009

Vessel, E. A., Starr, G. G., & Rubin, N. (2012). The brain on art: Intense aesthetic experience activates the default mode network. *Frontiers in Human Neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00066

Vessel, E. A., Starr, G. G., & Rubin, N. (2013). Art reaches within: Aesthetic experience, the self and the default mode network. *Frontiers in Human Neuroscience, 7*, Article 258. https://doi.org/10.3389/fnins.2013.00258

Wallot, S., & Menninghaus, W. (2018). Ambiguity effects of rhyme and meter. *Journal of Experimental Psychology: Learning, Memory, and Cognition., 44*(12), 1947-1954. https://doi.org/10.1037/xlm0000557

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

Woo, S. E., Chernyshenko, O. S., Longley, A., Zhang, Z. X., Chiu, C. Y., & Stark, S. E. (2014). Openness to experience: Its lower level structure, measurement, and cross-cultural equivalence. *Journal of Personality Assessment, 96*(1), 29-45. https://doi.org/10.1080/00223891.2013.806328

Table 1

Fit indices and test statistics for configural, metric and scalar invariance of the second-order factor model of the AReA between the US (n = 140) and German sample (n = 147) of random sample 2.

Fit index	Invariance test					
	Configural	Metric (first order factors)	Metric (second order factor)	Scalar (observed indicators)	Partial scalar [a] (observed indicators)	Partial Scalar [b] (first order factors)
χ^2	232.1	250.9	255.8	356.0	297.7	298.0
df	147	159	161	174	171	172
RMSEA	.064 [.048, .079]	.063 [.048, .078]	.064 [.049, .078]	.085 [.073, .098]	.072 [.058, .085]	.071 [.058, .085]
ΔRMSEA		-.001	.001	.021	.008	.007
CFI	.957	.954	.952	.908	.936	.937
ΔCFI		-.003	-.002	-.044	-.016	-.015
SRMR	.050	.069	.073	.089	.079	.080
ΔSRMR		.019	.004	.016	.006	.001

Note. The residual variance of one first-order factor in the German sample had a small negative estimate and was therefore set to zero in all models.

[a] Intercept equality constraints lifted for items 5, 11, and 12; test against metric (second order factor) invariance model.

[b] Equality constraints set for all first-order factor means and the second-order factor mean; test against partial scalar (observed indicators) invariance model. This final model is presented in Figure 1. See supplemental material for item wording.

Table 2

Mean scale scores, correlations with factor scores, and reliability estimates for AReA subscales and total score for the US (n = 281) and German sample (n = 293)

	US sample				German sample			
	AA	IAE	CB	AReA	AA	IAE	CB	AReA
Scale mean scores								
Mean	3.5	2.4	2.6	2.8	3.7	2.6	2.3	2.8
SD	0.8	0.9	1.0	0.8	0.7	0.8	1.0	0.7
S	-0.2	0.3	0.4	0.3	-0.5	0.4	0.7	0.3
K	2.6	2.5	2.6	2.6	3.3	2.9	2.6	3.2
p (SK)	.12	.007	.005	.057	.003	.060	.001	.12
r (scores)	.98	.98	.99	.90	.98	.98	.95	.97
Reliability (ω)	.91	.89	.72	.89	.86	.80	.73	.84

Note. AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score; SD = Standard deviation; S = Skewness; K = Kurtosis; *p* (SK) = Joint skewness/kurtosis test for normality; *r* (scores) = Pearson correlations of scale mean scores with factor scores. Tests of average differences in scale mean scores between the US and German samples showed that the US sample scored significantly lower on the AReA subscales AA, $t(572) = -3.4$, $p = .001$, and IAE, $t(572) = -2.5$, $p = .013$, but higher on CB, $t(572) = 3.8$, $p < .001$. In contrast, the AReA total score did not differ significantly between the samples, $t(572) = -0.4$, $p = .69$.

Table 3

Correlations between AReA subscales and total score and subscales of the TEPS (n = 50)

TEPS	AA	IAE	CB	AReA
TEPS-A	.38**	.15	.04	.18
TEPS-C	.44**	.37**	.24	.38**

** $p < .01$

Note. AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score; TEPS-A = Temporal Expectations of Pleasure Scale, Anticipatory Pleasure; TEPS-C = Temporal Expectations of Pleasure Scale, Consumatory Pleasure.

Table 4

Correlations between AReA subscales and total score and aesthetic judgments of visual paintings (n = 21).

Aesthetic judgments	AA	IAE	CB	AReA
1 second exposure				
Momentary force rating				
Mean	.10	.26	.22	.24
Maximum	.17	.35	.36	.36
Retrospective	-.09	.10	.20	.11
5 and 15 second exposure (combined)				
Momentary force rating				
Mean	.28	.44*	.35	.42*
Maximum	.28	.43*	.44*	.45*
Retrospective	.06	.28	.22	.23

* $p < .05$

Note. AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score. Momentary ratings are the average of the measured force produced during stimulus exposure. Retrospective ratings were provided on an analogue scale ranging from 0 to 1.

Table 5

Correlations between AReA subscales and total score and aesthetic judgments of auditory stimuli (n = 26)

Aesthtic judgments	AA	IAE	CB	AReA
Classical Music				
Momentary force rating				
Mean	.24	.31	.35	.35
Maximum	.44*	.31	.17	.31
Retrospective	.28	.31	.31	.34
Electronic Music				
Momentary force rating				
Mean	-.15	-.09	-.14	-.14
Maximum	.23	.13	-.16	.03
Retrospective	-.22	-.19	-.25	-.25

* $p < .05$

Note. AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score. Online ratings are the average of the measured force produced during stimulus exposure. Retrospective ratings were provided on an analogue scale ranging from 0 to 1.

Table 6

Correlations between AReA subscales and total score and average ratings of original poems, as well as differences in ratings for original vs. partly de-rhetorized poems (n = 40)

	AA	IAE	CB	AReA
Average ratings for original poems				
Beauty	.38*	.58***	.21	.47**
Movingness	.14	.36*	.32*	.34*
Melodiousness	.06	.31*	.16	.23
Joy	.10	.41**	-.001	.21
Sadness	.14	.34*	.32*	.33*
Absolute difference scores of original poems v. poem version without rhyme and meter				
Beauty	.24	.38*	.03	.26
Movingness	.24	.33*	.08	.26
Melodiousness	-.01	.22	.11	.14
Joy	.14	.40**	-.07	.19
Sadness	.12	.23	.02	.15

* $p < .05$, ** $p < .01$, *** $p < .001$.

Note. Ratings for beauty, movingness, and melodiousness were averaged across 10 poems, joy and sadness ratings only across the 5 joyful and sad poems from the same set; AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score.

Table 7

Correlations between AReA scale scores and subscales of BIS/BAS, BMRQ and chills (n = 167)

	AA	IAE	CB	AReA
BIS/BAS				
BIS total	-.01	.09	.03	.03
BAS total	.16*	.20**	.19*	.21**
BAS-Drive	.20**	.21**	.21**	.24**
BAS-Fun-Seeking	.25**	.31***	.14	.27***
BAS-Reward	.25***	.29***	.22**	.30***
BMRQ				
Music Seeking	.39***	.26***	.20*	.35***
Emotional Evocation	.36***	.25**	.11	.30***
Mood Regulation	.32***	.14	.08	.25**
Sensory-Motor	.14	.10	.03	.12
Social Reward	.39***	.23**	.15	.33***
Chills				
Trait	.16*	.25**	.18*	.24**
State	.24**	.26**	.09	.25**

* $p < .05$, ** $p < .01$, *** $p < .001$.

Note. AA = Aesthetic Appreciation; BIS/BAS = Behavioral Inhibition/Activation System; BMRQ = Barcelona Music Reward Questionnaire; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score.

Table 8

Correlations between AReA subscales and total score and Big Five Inventory 2 domain scales and facet scales of the domain Open-Mindedness (n = 207)

		Correlations with AReA scales			
	Mean (SD)	AA	IAE	CB	AReA
BFI-2 domains					
Extraversion	40.5 (7.3)	.30***	.21**	.17*	.29***
Agreeableness	45.2 (6.0)	.18*	.13	.07	.16*
Conscientiousness	43.5 (7.2)	.12	-.02	-.01	.06
Negative Emotionality	32.1 (7.7)	.03	.12	.06	.07
Open-Mindedness	47.1 (7.0)	.61***	.45***	.48***	.63***
BFI-2 facets of Open-Mindedness					
Intellectual Curiosity	15.9 (2.8)	.35***	.27***	.28***	.37***
Aesthetic Sensitivity	16.5 (2.9)	.71***	.42***	.26***	.64***
Creative Imagination	14.7 (3.4)	.36***	.35***	.44***	.45***

* $p < .05$, ** $p < .01$, *** $p < .001$.

Note. AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score.

Figure 1

Final CFA model for the AReA in the English and German language version including unstandardized coefficients from the partial scalar invariance model. First- and second-order factor loading parameters are equal for the two version. Residual variances of first-order factors and the variance of the second-order factor shown are for the English version in the first line and for the German version in the second line. Item intercepts and error variances not shown.