

Goldsmiths Research Online

*Goldsmiths Research Online (GRO)
is the institutional research repository for
Goldsmiths, University of London*

Citation

Schlotz, Wolff; Wallot, Sebastian; Omigie, Diana; Masucci4, Michael D.; Hoelzmann, Sonja C. and Vessel, Edward A.. 2021. The Aesthetic Responsiveness Assessment (AReA): A screening tool to assess individual differences in responsiveness to art in English and German. *Psychology of Aesthetics, Creativity, and the Arts*, 15(4), pp. 682-696. ISSN 1931-3896 [Article]

Persistent URL

<https://research.gold.ac.uk/id/eprint/29310/>

Versions

The version presented here may differ from the published, performed or presented work. Please go to the persistent GRO record above for more information.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Goldsmiths, University of London via the following email address: gro@gold.ac.uk.

The item will be removed from the repository while any claim is being investigated. For more information, please contact the GRO team: gro@gold.ac.uk

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

3 Wolff Schlotz* ^{1,2}, Sebastian Wallot* ¹, Diana Omigie^{1,3}, Michael D. Masucci⁴, Sonja C.

4 Hoelzmann^{1,2}, and Edward A. Vessel¹

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

6 ² Institute of Psychology, Goethe-University Frankfurt am Main, Germany

7 ³ Department of Psychology, Goldsmiths, University of London, UK

8 ⁴ Department of Psychology, New York University

9 **Author Note**

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

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

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

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

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

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

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

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

18 We have no known conflict of interest to disclose.

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

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

26 **Abstract**

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

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

45

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

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

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

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

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

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

88 The construct of aesthetic responsiveness is related to constructs focusing on
89 individual differences in the appreciation of, or engagement with beauty (Diessner et al.,
90 2018; Diessner et al., 2008; Haidt & Keltner, 2004), particularly if appreciation is conceived
91 as a cognitive-emotional, and engagement as an emotional reaction to beauty (Güsewell &
92 Ruch, 2012). However, aesthetic responsiveness differs from these constructs in a number of
93 aspects. First, it focuses on responses to aesthetic stimuli and excludes non-aesthetic stimuli
94 such as talent, virtue, or morality. Second, it explicitly distinguishes between response
95 domains, providing a background for more fine-grained predictions of domain-specific

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

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

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

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

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

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

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

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

159 **Scale Construction**

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

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

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

174 Another key goal of scale construction was to distinguish between those individuals
175 who regularly respond to artworks in an intense way from those who rarely experience more
176 than a commonplace appreciation of aesthetic objects in everyday life. Recent empirical work
177 suggests a potential difference between more everyday positive experiences of beauty and a
178 subset of more intense aesthetic experiences (e.g. “being moved”, “awe”, the “sublime”;
179 (Briellmann & Pelli, 2017; Omigie et al., 2019; Pelowski et al., 2017; Vessel et al., 2012,
180 2013).

181 Such work parallels accounts in the philosophical literature that pit feelings of beauty
182 against those of the sublime (Burke, 1757/2015). In the context of music, for instance, beauty
183 experiences “in which tension and discord have at most a minor place” have been
184 distinguished from other forms of beauty, that may, instead, confront or challenge (Levinson,
185 2012, p. 128). Here, we sought to extend, to the individual differences level, this notion of a
186 distinction in the types of aesthetic states that are possible. We propose that a scale that is able
187 to reveal those individuals that regularly respond to artworks in an intense way would allow
188 experimenters to better account for much variability in responses observable in their data.

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

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

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

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

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

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

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

252 Method

253 **Samples**

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

264 **German sample.** The German sample consisted of two subsamples. German
265 subsample 1 was a convenience sample of participants from a study on music listening
266 behavior. For this study, 202 participants were recruited, of which 31 did not provide any
267 responses on the aesthetic responsiveness scale, and one had 78% missing responses.
268 Removing these participants resulted in a final sample of 170 participants, 118 females (69%)
269 (7 participants, 4%, did not respond), with a mean age of 31.1 years ($SD = 12.5$; range: 18 to
270 75 years); 73 (43 %) had completed a university degree.

271 German subsample 2 was a convenience sample from a study of poem reading. After
272 the reading study, participants filled in the aesthetic responsiveness scale as part of a larger set
273 of questions. The sample consisted of 123 participants, 92 (75%) females, with a mean age of
274 25.0 years ($SD = 5.1$; range: 18 to 43 years); 54 (44 %) completed a university degree.

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

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

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

282 **Measures**

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

287 **Data analysis**

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

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

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

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

331 Although we report chi-square differences ($\Delta\chi^2$) for all model comparisons, our
332 decisions on measurement invariance were based on differences in approximate fit indices, as
333 $\Delta\chi^2$ is highly sensitive to sample size. In particular, differences in CFI (ΔCFI), RMSEA
334 (ΔRMSEA), and SRMR (ΔSRMR) between models with increasing restrictions were used to
335 assess each level of measurement invariance. In the case of metric invariance, changes of
336 $\Delta\text{CFI} \leq -.010$, $\Delta\text{RMSEA} \geq .015$, and $\Delta\text{SRMR} \geq .015$ would indicate non-invariance as
337 suggested by Cheung and Rensvold (2002) and Chen (2007). In the case of scalar invariance,
338 $\Delta\text{SRMR} \geq .010$ would indicate non-invariance, with the other criteria being the same as for
339 metric invariance, as suggested by Chen (2007).

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

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

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

357 Results

358 Item selection and factor analyses

359 Although the items were designed to indicate different facets of a disposition to
360 respond to aesthetic stimuli, we assumed that they share variance attributable to a common
361 underlying factor, i.e. aesthetic responsiveness. We therefore expected all items to show
362 relatively high associations with the scale score minus the item itself, i.e. IRC, and at least
363 medium inter-item correlations. Sample-specific IRCs as well as average inter-item
364 correlations were higher in the English language version than in the German language version
365 (see Tables S1 and S2 in the supplementary material for details); three items showed very
366 weak IRCs of less than .30 in the German language version, one of which was also very weak
367 in the English language version. We therefore excluded these items (number 7, 15, and 17 of
368 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.

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

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

386 To check stability of the factorial structure across random samples, we conducted
387 second-order CFAs using random sample 2. CFA models were fitted separately for the
388 English and German language versions. The CFA model showed an acceptable fit to the data
389 in both, the English language ($\chi^2 = 112.6$; $df = 73$; $p = .002$; RMSEA = 0.062, 90% CI: 0.038,
390 0.084; CFI = 0.965; TLI = 0.957; SRMR = 0.050) and German language version ($\chi^2 = 119.6$;
391 $df = 74$ (the residual variance of one first-order factor in the German sample had a small
392 negative estimate and was therefore set to zero); $p = .001$; RMSEA = 0.065, 90% CI: 0.042,
393 0.086; CFI = 0.946; TLI = 0.933; SRMR = 0.050). These results provide support for the

394 validity of the factorial structure across different samples.

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

402 **Measurement invariance across language versions**

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

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

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

428 Fitting the CFA model shown in Figure 1 to data from another German validation
429 sample of 207 participants resulted in a good model fit ($\chi^2 = 110.1$; $df = 73$; $p = .003$; RMSEA
430 = 0.050, 90% CI: 0.029, 0.068; CFI = 0.958; TLI = 0.948; SRMR = 0.052). Factor loadings
431 and latent factor correlations (not shown here) were similar to the results for random sample 2
432 shown in Figure 1. These results further support the factorial validity of the AReA German
433 language version.

434 **Scale scores**

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

443 **Reliability**

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

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

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

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

463 For the TEPS, we expected both scales to show a positive relationship to the AReA
464 sub-scales AA and IAE. Specifically, the TEPS-C scale should bear a positive relationship
465 with the AReA sub-scales, because openness to experience is conceptually closely linked with
466 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.

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

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

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

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

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

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

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

512 **Validation study 2: Responses to poems**

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

517 rhetorized version. Additionally, all participants filled in the AReA and provided ratings of
518 different versions of poems on a 7-point scale for beauty, movingness, melodiousness, joy,
519 and sadness. Previous research on poem and proverb reading has shown that manipulations of
520 rhyme and meter lead to changes in the processing and aesthetic evaluation of language
521 (Menninghaus, Bohrn, et al., 2015; Menninghaus & Wallot, 2020; Wallot & Menninghaus,
522 2018).

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

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

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

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

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

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

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

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

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

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

598 In another German validation sample, an online survey presented the final 14-item
599 AReA version as well as a German translation of the BFI-2 (Danner et al., 2019; Soto & John,
600 2017) and was completed by 207 participants (3 participants were excluded due to extremely
601 long response times). We computed Pearson's correlation coefficients between AReA scale
602 scores and the BFI-2 domain scales as well as the three facet scales constituting Open-
603 Mindedness, i.e. Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination. The
604 pattern of correlations will provide additional information on the convergent and discriminant
605 validity of the AReA scales. We expected large correlations between AReA scales and the
606 Open-Mindedness scale, but much smaller correlations with the other domain scales, i.e.
607 Extraversion, Agreeableness, Conscientiousness, and Negative Emotionality. With regard to
608 the facet scales of Open-Mindedness, large correlations with AReA scales were expected for
609 the facet Aesthetic Sensitivity, whereas correlations with the other facet scales were expected
610 to be much smaller. Finally, the correlation between the AReA subscale Creative Behavior
611 and the facet scale Creative Imagination was expected to be higher than with the other facet
612 scales, as an individual disposition to high levels of creative imagination is expected to
613 facilitate creative behavior as assessed by the AReA subscale.

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

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

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

630 Discussion

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

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

642 between responses to nature, art and moral beauty).

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

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

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

671 Using independent samples or sub-samples of participants that took part in different
672 studies on the reception and evaluation of music, visual art, and poetry, we found evidence
673 supporting the validity of scale scores by showing expected correlations with self reported
674 strength of aesthetic responsiveness to visual (validation study 1), musical (validation studies
675 1 and 2) and literary aesthetic stimuli (validation study 3), as well as scales tapping into
676 general (BIS/BAS and TEPS), and more domain-specific hedonic responses (BMRQ).
677 Although due to small sample sizes not all of these correlations were statistically significant,
678 many of them represent rather large effects from a normative perspective (Gignac & Szodorai,
679 2016). These results suggest a broad applicability of AReA as a screening instrument across a
680 variety of domains of art perception.

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

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

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

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

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

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

733 It is important to note that the construct of aesthetic responsiveness explicitly excludes
734 reference to an external standard and is therefore very different from constructs that refer to
735 quality of judgements of aesthetic stimuli such as aesthetic sensitivity (Child, 1964; Eysenck,
736 1940; Myszkowski & Zenasni, 2016; but see Corradi et al., 2019). This has the great
737 advantage that the AReA can be used in non-experts and experts alike. Our theoretical
738 approach clearly implies that the question of whether these groups differ in their aesthetic
739 responsiveness is not a theoretical but an empirical issue. However, the construct defined here
740 nevertheless refers to responsiveness to aesthetic stimuli, and any measure of the construct
741 has to demonstrate that scores reflect more than just non-specific responsivity. In this sense,

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

745 **Limitations and outlook**

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

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

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

774 **Conclusion**

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

785 **References**

- 786 Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the
787 HEXACO model of personality structure. *Personality and Social Psychology Review*,
788 *11*(2), 150-166. <https://doi.org/10.1177/1088868306294907>
- 789 Belfi, A. M., Vessel, E. A., Briemann, A., Isik, A. I., Chatterjee, A., Leder, H., Pelli, D. G.,
790 & Starr, G. G. (2019). Dynamics of aesthetic experience are reflected in the default-
791 mode network. *Neuroimage*, *188*, 584-597.
792 <https://doi.org/10.1016/j.neuroimage.2018.12.017>
- 793 Berlyne, D. E. (1971). *Aesthetics and psychobiology*. Appleton-Century-Crofts.
- 794 Briemann, A. A., & Pelli, D. G. (2017). Beauty requires thought. *Current Biology*, *27*(10),
795 1506-1513.e3. <https://doi.org/10.1016/j.cub.2017.04.018>
- 796 Burke, E. (2015). *A philosophical enquiry into the sublime and beautiful*. Oxford University
797 Press. (Original work published 1757)
- 798 Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and
799 affective responses to impending reward and punishment: The BIS/BAS scales.
800 *Journal of Personality and Social Psychology*, *67*, 319-333.
801 <https://doi.org/10.1037/0022-3514.67.2.319>
- 802 Chatterjee, A., & Vartanian, O. (2016). Neuroscience of aesthetics. *Annals of the New York*
803 *Academy of Sciences*, *1369*(1), 172-194. <https://doi.org/10.1111/nyas.13035>
- 804 Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance.
805 *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504.
806 <https://doi.org/10.1080/10705510701301834>
- 807 Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-
808 order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*,
809 *12*(3), 471-492. https://doi.org/10.1207/s15328007sem1203_7

- 810 Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing
811 measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*,
812 9(2), 233-255. https://doi.org/10.1207/s15328007sem0902_5
- 813 Child, I. L. (1964). Observations on the meaning of some measures of esthetic sensitivity.
814 *Journal of Psychology*, 57(1), 49-64. <https://doi.org/10.1080/00223980.1964.9916671>
- 815 Corradi, G., Chuquichambi, E. G., Barrada, J. R., Clemente, A., & Nadal, M. (2019). A new
816 conception of visual aesthetic sensitivity. Advance online publication. *British Journal*
817 *of Psychology*. <https://doi.org/10.1111/bjop.12427>
- 818 Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality
819 assessment using the Revised NEO Personality Inventory. *Journal of Personality*
820 *Assessment*, 64(1), 21-50. https://doi.org/10.1207/s15327752jpa6401_2
- 821 Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., &
822 John, O. P. (2019). Das Big Five Inventar 2 [The Big Five Inventory 2]. *Diagnostica*,
823 65(3), 121-132. <https://doi.org/10.1026/0012-1924/a000218>
- 824 Diessner, R., Pohling, R., Stacy, S., & Güsewell, A. (2018). Trait appreciation of beauty: A
825 story of love, transcendence, and inquiry. *Review of General Psychology*, 22(4), 377-
826 397. <https://doi.org/10.1037/gpr0000166>
- 827 Diessner, R., Solom, R. C., Frost, N. K., Parsons, L., & Davidson, J. (2008). Engagement with
828 beauty: appreciating natural, artistic, and moral beauty. *Journal of Psychology*, 142(3),
829 303-329. <https://doi.org/10.3200/jrlp.142.3.303-332>
- 830 Elvers, P., Omigie, D., Fuhrmann, W., & Fischinger, T. (2015). Exploring the musical taste of
831 expert listeners: musicology students reveal tendency toward omnivorous taste.
832 *Frontiers in Psychology*, 6, Article 1252. <https://doi.org/10.3389/fpsyg.2015.01252>
- 833 Eysenck, H. J. (1940). The general factor in aesthetic judgement. *British Journal of*
834 *Psychology*, 31(1), 94-102. <https://doi.org/10.1111/j.2044-8295.1940.tb00977.x>

- 835 Fayn, K., MacCann, C., Tiliopoulos, N., & Silvia, P. J. (2015). Aesthetic emotions and
836 aesthetic people: Openness predicts sensitivity to novelty in the experiences of interest
837 and pleasure. *Frontiers in Psychology, 6*, Article 1877.
838 <https://doi.org/10.3389/fpsyg.2015.01877>
- 839 Gard, D. E., Gard, M. G., Kring, A. M., & John, O. P. (2006). Anticipatory and
840 consummatory components of the experience of pleasure: A scale development study.
841 *Journal of Research in Personality, 40*(6), 1086-1102.
842 <https://doi.org/10.1016/j.jrp.2005.11.001>
- 843 Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences
844 researchers. *Personality and Individual Differences, 102*, 74-78.
845 <https://doi.org/10.1016/j.paid.2016.06.069>
- 846 Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new visual aesthetic
847 sensitivity test I: Construction and psychometric properties. *Perceptual and Motor*
848 *Skills, 49*(3), 795-802. <https://doi.org/10.2466/pms.1979.49.3.795>
- 849 Güsewell, A., & Ruch, W. (2012). Are there multiple channels through which we connect
850 with beauty and excellence? *Journal of Positive Psychology, 7*(6), 516-529.
851 <https://doi.org/10.1080/17439760.2012.726636>
- 852 Hager, M., Hagemann, D., Danner, D., & Schankin, A. (2012). Assessing aesthetic
853 appreciation of visual artworks—The construction of the Art Reception Survey
854 (ARS). *Psychology of Aesthetics, Creativity, and the Arts, 6*(4), 320-333.
855 <https://doi.org/10.1037/a0028776>
- 856 Haidt, J., & Keltner, D. (2004). Appreciation of beauty and excellence [awe, wonder,
857 elevation]. In C. K. Peterson & M. E. P. Seligman (Eds.), *Character strengths and*
858 *virtues* (pp. 537-551). Oxford University Press.
- 859 Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure

- 860 analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling:*
861 *A Multidisciplinary Journal*, 6, 1-55.
- 862 Istok, E., Brattico, E., Jacobsen, T., Krohn, K., Muller, M., & Tervaniemi, M. (2009).
863 Aesthetic responses to music: A questionnaire study. *Musicae Scientiae*, 13(2), 183-
864 206. <https://doi.org/10.1177/102986490901300201>
- 865 Jacobsen, T., & Beudt, S. (2017). Domain generality and domain specificity in aesthetic
866 appreciation. *New Ideas in Psychology*, 47, 97-102.
867 <https://doi.org/10.1016/j.newideapsych.2017.03.008>
- 868 Jacobsen, T., Buchta, K., Köhler, M., & Schröger, E. (2004). The primacy of beauty in
869 judging the aesthetics of objects. *Psychological Reports*, 94(3_suppl), 1253-1260.
870 <https://doi.org/10.2466/pr0.94.3c.1253-1260>
- 871 Karwowski, M., & Beghetto, R. A. (2019). Creative behavior as agentic action. *Psychology of*
872 *Aesthetics, Creativity, and the Arts*, 13(4), 402-415.
873 <https://doi.org/10.1037/aca0000190>
- 874 Lane, R. D., Waldstein, S. R., Chesney, M. A., Jennings, J. R., Lovallo, W. R., Kozel, P. J.,
875 Rose, R. M., Drossman, D. A., Schneiderman, N., Thayer, J. F., & Cameron, O. G.
876 (2009). The rebirth of neuroscience in psychosomatic medicine, part I: Historical
877 context, methods, and relevant basic science. *Psychosomatic Medicine*, 71(2), 117-
878 134. <https://doi.org/10.1097/PSY.0b013e31819783be>
- 879 Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion
880 and evaluation in art appreciation. *Cognition and Emotion*, 28(6), 1137-1147.
881 <https://doi.org/10.1080/02699931.2013.870132>
- 882 Levinson, J. (2012). Musical beauty. *Teorema*, 31(3), 127-135.
- 883 Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its
884 recent variants. *Psychological Methods*, 24(4), 452-467.

- 885 <https://doi.org/10.1037/met0000230>
- 886 Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells,
887 A. (2013). Individual differences in music reward experiences. *Music Perception*,
888 *31*(2), 118-138. <https://doi.org/10.1525/mp.2013.31.2.118>
- 889 McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience.
890 *Motivation and Emotion*, *31*(1), 5-11. <https://doi.org/10.1007/s11031-007-9053-1>
- 891 McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.
- 892 McManus, I. C., & Furnham, A. (2006). Aesthetic activities and aesthetic attitudes: Influences
893 of education, background and personality on interest and involvement in the arts.
894 *British Journal of Psychology*, *97*, 555-587.
895 <https://doi.org/10.1348/000712606x101088>
- 896 McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological*
897 *Methods*, *23*(3), 412-433. <https://doi.org/10.1037/met0000144>
- 898 Menninghaus, W., Bohrn, I. C., Knoop, C. A., Kotz, S. A., Schlotz, W., & Jacobs, A. M.
899 (2015). Rhetorical features facilitate prosodic processing while handicapping ease of
900 semantic comprehension. *Cognition*, *143*, 48-60.
901 <https://doi.org/10.1016/j.cognition.2015.05.026>
- 902 Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T., & Koelsch, S.
903 (2017). The Distancing-Embracing Model of the enjoyment of negative emotions in
904 art reception. *Behavioral and Brain Sciences*, *40*, Article e347.
905 <https://doi.org/10.1017/S0140525X17000309>
- 906 Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T.
907 (2015). Towards a psychological construct of being moved. *PLOS ONE*, *10*(6), Article
908 e0128451. <https://doi.org/10.1371/journal.pone.0128451>
- 909 Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., & Knoop, C. A. (2017). The

- 910 emotional and aesthetic powers of parallelistic diction. *Poetics*, 63, 47-59.
911 <https://doi.org/10.1016/j.poetic.2016.12.001>
- 912 Menninghaus, W., Wagner, V., Wassiliwizky, E., Schindler, I., Hanich, J., Jacobsen, T., &
913 Koelsch, S. (2019). What are aesthetic emotions? *Psychological Review*, 126(2), 171-
914 195. <https://doi.org/10.1037/rev0000135>
- 915 Menninghaus, W., & Wallot, S. (2020). *What the eyes reveal about (reading) poetry*
916 [Manuscript submitted for publication]. Department of Language and Literature, Max
917 Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany.
- 918 Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Psychology Press.
- 919 Myszkowski, N., Storme, M., Zenasni, F., & Lubart, T. (2014). Is visual aesthetic sensitivity
920 independent from intelligence, personality and creativity? *Personality and Individual*
921 *Differences*, 59(0), 16-20. <https://doi.org/10.1016/j.paid.2013.10.021>
- 922 Myszkowski, N., & Zenasni, F. (2016). Individual Differences in Aesthetic Ability: The Case
923 for an Aesthetic Quotient. *Frontiers in Psychology*, 7, Article 750.
924 <https://doi.org/10.3389/fpsyg.2016.00750>
- 925 Omigie, D., Frieler, K., Bär, C., Muralikrishnan, R., Wald-Fuhrmann, M., & Fischinger, T.
926 (2019). Experiencing musical beauty: Emotional subtypes and their physiological and
927 musico-acoustic correlates. *Psychology of Aesthetics, Creativity, and the Arts*,
928 Advance online publication. <https://doi.org/10.1037/aca0000271>
- 929 Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish
930 me... delight my eyes and brain: The Vienna Integrated Model of top-down and
931 bottom-up processes in Art Perception (VIMAP) and corresponding affective,
932 evaluative, and neurophysiological correlates. *Physics of Life Reviews*, 21, 80-125.
933 <https://doi.org/10.1016/j.plrev.2017.02.003>
- 934 Peterson, C. K., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook*

- 935 *and classification*. American Psychological Association.
- 936 Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential
937 conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity,*
938 *and the Arts, 11*(1), 59-68. <https://doi.org/10.1037/aca0000079>
- 939 Rawlings, D., Barrantes i Vidal, N., & Furnham, A. (2000). Personality and aesthetic
940 preference in Spain and England: two studies relating sensation seeking and openness
941 to experience to liking for paintings and music. *European Journal of Personality,*
942 *14*(6), 553-576. [https://doi.org/10.1002/1099-0984\(200011/12\)14:6<553::AID-](https://doi.org/10.1002/1099-0984(200011/12)14:6<553::AID-PER384>3.0.CO;2-H)
943 PER384>3.0.CO;2-H
- 944 Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure:
945 is beauty in the perceiver's processing experience? *Personality and Social*
946 *Psychology Review, 8*(4), 364-382. https://doi.org/10.1207/s15327957pspr0804_3
- 947 Rowold, J. (2008). Instrument development for esthetic perception assessment. *Journal of*
948 *Media Psychology, 20*(1), 35-40. <https://doi.org/10.1027/1864-1105.20.1.35>
- 949 Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., & Scherer,
950 K. R. (2017). Measuring aesthetic emotions: A review of the literature and a new
951 assessment tool. *PLOS ONE, 12*(6), Article e0178899.
952 <https://doi.org/10.1371/journal.pone.0178899>
- 953 Schlotz, W. (2013). Stress reactivity. In M. D. Gellman & R. J. Turner (Eds.), *Encyclopedia*
954 *of Behavioral Medicine* (pp. 1891-1894). Springer.
- 955 Schlotz, W., Yim, I. S., Zoccola, P. M., Jansen, L., & Schulz, P. (2011). The perceived stress
956 reactivity scale: Measurement invariance, stability, and validity in three countries.
957 *Psychological Assessment, 23*(1), 80-94. <https://doi.org/10.1037/a0021148>
- 958 Silvia, P. J., Fayn, K., Nusbaum, E. C., & Beaty, R. E. (2015). Openness to experience and
959 awe in response to nature and music: Personality and profound aesthetic experiences.

- 960 *Psychology of Aesthetics, Creativity, and the Arts*, 9(4), 376-384.
- 961 <https://doi.org/10.1037/aca0000028>
- 962 Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and
963 assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and
964 predictive power. *Journal of Personality and Social Psychology*, 113(1), 117-143.
965 <https://doi.org/10.1037/pspp0000096>
- 966 Stamatopoulou, D. (2004). Integrating the philosophy and psychology of aesthetic experience:
967 Development of the aesthetic experience scale. *Psychological Reports*, 95(2), 673-
968 695. <https://doi.org/10.2466/PRO.95.6.673-695>
- 969 Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial
970 measurement invariance enough? *Methodology*, 9(1), 1-12.
971 <https://doi.org/10.1027/1614-2241/a000049>
- 972 Strobel, A., Beauducel, A., Debener, S., & Brocke, B. (2001). Eine deutschsprachige Version
973 des BIS/BAS-Fragebogens von Carver und White [A German language version of the
974 BIS/BAS questionnaire by Carver and White]. *Zeitschrift für Differentielle und*
975 *Diagnostische Psychologie*, 22(3), 216-227. [https://doi.org/10.1024//0170-
976 1789.22.3.216](https://doi.org/10.1024//0170-1789.22.3.216)
- 977 van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines.
978 *European Psychologist*, 1(2), 89-99. <https://doi.org/10.1027/1016-9040.1.2.89>
- 979 Vessel, E. A., Isik, A. I., Belfi, A. M., Stahl, J. L., & Starr, G. G. (2019). The default-mode
980 network represents aesthetic appeal that generalizes across visual domains.
981 *Proceedings of the National Academy of Sciences*, 116(38), 19155-19164.
982 <https://doi.org/10.1073/pnas.1902650116>
- 983 Vessel, E. A., Maurer, N., Denker, A. H., & Starr, G. G. (2018). Stronger shared taste for
984 natural aesthetic domains than for artifacts of human culture. *Cognition*, 179, 121-131.

- 985 <https://doi.org/10.1016/j.cognition.2018.06.009>
- 986 Vessel, E. A., Starr, G. G., & Rubin, N. (2012). The brain on art: Intense aesthetic experience
987 activates the default mode network. *Frontiers in Human Neuroscience*, 6.
988 <https://doi.org/10.3389/fnhum.2012.00066>
- 989 Vessel, E. A., Starr, G. G., & Rubin, N. (2013). Art reaches within: Aesthetic experience, the
990 self and the default mode network. *Frontiers in Human Neuroscience*, 7, Article 258.
991 <https://doi.org/10.3389/fnins.2013.00258>
- 992 Wallot, S., & Menninghaus, W. (2018). Ambiguity effects of rhyme and meter. *Journal of*
993 *Experimental Psychology: Learning, Memory, and Cognition.*, 44(12), 1947-1954.
994 <https://doi.org/10.1037/xlm0000557>
- 995 Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T., & Menninghaus, W. (2017). The
996 emotional power of poetry: neural circuitry, psychophysiology and compositional
997 principles. *Social, Cognitive, and Affective Neuroscience*, 12(8), 1229-1240.
998 <https://doi.org/10.1093/scan/nsx069>
- 999 Woo, S. E., Chernyshenko, O. S., Longley, A., Zhang, Z. X., Chiu, C. Y., & Stark, S. E.
1000 (2014). Openness to experience: Its lower level structure, measurement, and cross-
1001 cultural equivalence. *Journal of Personality Assessment*, 96(1), 29-45.
1002 <https://doi.org/10.1080/00223891.2013.806328>
1003

1004 **Table 1**

1005 *Fit indices and test statistics for configural, metric and scalar invariance of the second-order*
 1006 *factor model of the AReA between the US (n = 140) and German sample (n = 147) of random*
 1007 *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]
Δ RMSEA		-.001	.001	.021	.008	.007
CFI	.957	.954	.952	.908	.936	.937
Δ CFI		-.003	-.002	-.044	-.016	-.015
SRMR	.050	.069	.073	.089	.079	.080
Δ SRMR		.019	.004	.016	.006	.001

1008 Note. The residual variance of one first-order factor in the German sample had a small
 1009 negative estimate and was therefore set to zero in all models.

1010 ^a Intercept equality constraints lifted for items 5, 11, and 12; test against metric (second order
 1011 factor) invariance model.

1012 ^b Equality constraints set for all first-order factor means and the second-order factor mean;
 1013 test against partial scalar (observed indicators) invariance model. This final model is
 1014 presented in Figure 1. See supplemental material for item wording.

1015 **Table 2**1016 *Mean scale scores, correlations with factor scores, and reliability estimates for AReA*1017 *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
<i>p</i> (SK)	.12	.007	.005	.057	.003	.060	.001	.12
<i>r</i> (scores)	.98	.98	.99	.90	.98	.98	.95	.97
Reliability (ω)	.91	.89	.72	.89	.86	.80	.73	.84

1018 *Note.* AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative

1019 Behavior; AReA = Aesthetic Responsiveness Assessment total score; SD = Standard

1020 deviation; S = Skewness; K = Kurtosis; *p* (SK) = Joint skewness/kurtosis test for normality; *r*

1021 (scores) = Pearson correlations of scale mean scores with factor scores. Tests of average

1022 differences in scale mean scores between the US and German samples showed that the US

1023 sample scored significantly lower on the AReA subscales AA, $t(572) = -3.4$, $p = .001$, and1024 IAE, $t(572) = -2.5$, $p = .013$, but higher on CB, $t(572) = 3.8$, $p < .001$. In contrast, the AReA1025 total score did not differ significantly between the samples, $t(572) = -0.4$, $p = .69$.

1026

1027 **Table 3**1028 *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**

1029 ** $p < .01$ 1030 *Note.* AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative

1031 Behavior; AReA = Aesthetic Responsiveness Assessment total score; TEPS-A = Temporal

1032 Expectations of Pleasure Scale, Anticipatory Pleasure; TEPS-C = Temporal Expectations of

1033 Pleasure Scale, Consumatory Pleasure.

1034

1035 **Table 4**1036 *Correlations between AReA subscales and total score and aesthetic judgments of visual*1037 *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

1038 * $p < .05$ 1039 *Note.* AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative

1040 Behavior; AReA = Aesthetic Responsiveness Assessment total score. Momentary ratings are

1041 the average of the measured force produced during stimulus exposure. Retrospective ratings

1042 were provided on an analogue scale ranging from 0 to 1.

1043

1044 **Table 5**1045 *Correlations between AReA subscales and total score and aesthetic judgments of auditory*1046 *stimuli (n = 26)*

Aesthetic 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

1047 * $p < .05$ 1048 *Note.* AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative

1049 Behavior; AReA = Aesthetic Responsiveness Assessment total score. Online ratings are the

1050 average of the measured force produced during stimulus exposure. Retrospective ratings were

1051 provided on an analogue scale ranging from 0 to 1.

1052

1053 **Table 6**

1054 *Correlations between AReA subscales and total score and average ratings of original poems,*
 1055 *as well as differences in ratings for original vs. partly de-rhetorized poems (n = 40)*

	AA	IAE	CB	AReA
<i>Average ratings for original poems</i>				
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*
<i>Absolute difference scores of original poems v. poem version without rhyme and meter</i>				
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

1056 * $p < .05$, ** $p < .01$, *** $p < .001$.

1057 *Note.* Ratings for beauty, movingness, and melodiousness were averaged across 10 poems,

1058 joy and sadness ratings only across the 5 joyful and sad poems from the same set; AA =

1059 Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative Behavior; AReA

1060 = Aesthetic Responsiveness Assessment total score.

1061

1062 **Table 7**1063 *Correlations between AReA scale scores and subscales of BIS/BAS, BMRQ and chills (n =*1064 *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**

1065 * $p < .05$, ** $p < .01$, *** $p < .001$.1066 *Note.* AA = Aesthetic Appreciation; BIS/BAS = Behavioral Inhibition/Activation System;

1067 BMRQ = Barcelona Music Reward Questionnaire; IAE = Intense Aesthetic Experience; CB =

1068 Creative Behavior; AReA = Aesthetic Responsiveness Assessment total score.

1069

1070 **Table 8**1071 *Correlations between AReA subscales and total score and Big Five Inventory 2 domain scales*1072 *and facet scales of the domain Open-Mindedness (n = 207)*

	Mean (SD)	Correlations with AReA scales			
		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***

1073 * $p < .05$, ** $p < .01$, *** $p < .001$.1074 *Note.* AA = Aesthetic Appreciation; IAE = Intense Aesthetic Experience; CB = Creative

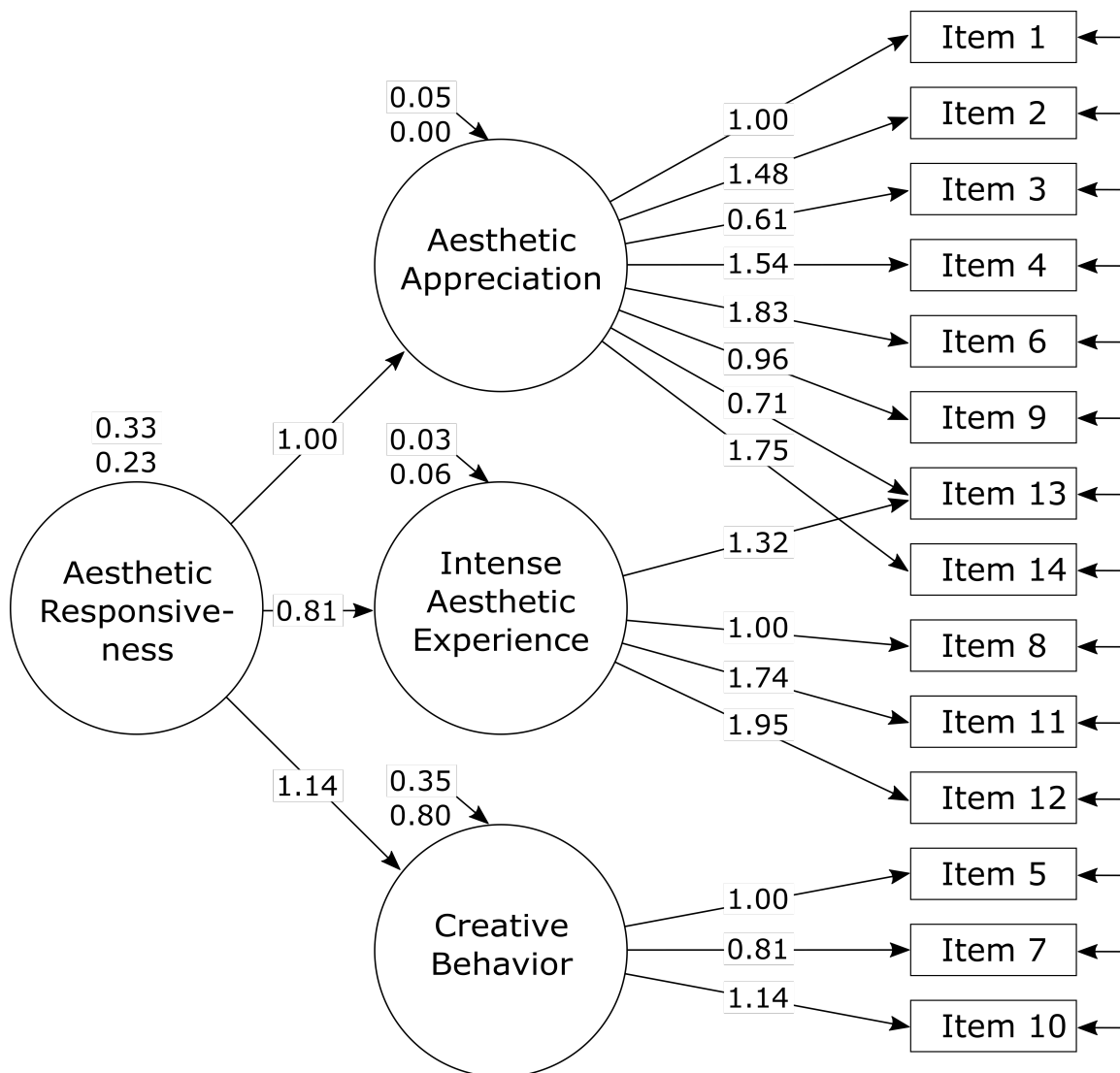
1075 Behavior; AReA = Aesthetic Responsiveness Assessment total score.

1076

1077

1078 **Figure 1**

1079 Final CFA model for the AReA in the English and German language version including
 1080 unstandardized coefficients from the partial scalar invariance model. First- and second-order
 1081 factor loading parameters are equal for the two version. Residual variances of first-order
 1082 factors and the variance of the second-order factor shown are for the English version in the
 1083 first line and for the German version in the second line. Item intercepts and error variances
 1084 not shown.



1085