Measuring Art Knowledge:
Item Response Theory Analysis of the Aesthetic Fluency Scale

Katherine N. Cotter[1], David F. Chen[2], Alexander P. Christensen[1],
Kyung Yong Kim[2], and Paul J. Silvia[2]
[1]University of Pennsylvania
[2]University of North Carolina at Greensboro

AUTHOR NOTE

Katherine N. Cotter, Positive Psychology Center, University of Pennsylvania; David F. Chen and Kyung Yong Kim, Department of Educational Research Methodology, University of North Carolina at Greensboro; Alexander P. Christensen, Penn Center for Neuroaesthetics, Department of Neurology, University of Pennsylvania; Paul J. Silvia, Department of Psychology, University of North Carolina at Greensboro.

Thank you to Jeff Smith, Lisa Smith, and Eva Specker for sharing data for these analyses.

Correspondence regarding this article should be addressed to Katherine N. Cotter, Positive Psychology Center, University of Pennsylvania, Philadelphia, PA; email: katherinencotter@gmail.com.

Abstract

The Aesthetic Fluency Scale is a commonly used measure of people's art knowledge. This scale was initially developed for museum visitors, but its usage has expanded to other populations, including non-arts students. The present research used an Item Response Theory approach to better understand its psychometric properties in two samples—artistically engaged individuals (i.e., museum visitors and art students) and non-arts students. Overall, terms related to art styles were easiest, the non-arts students found the scale more difficult than the artistically engaged, and four items (John Singer Sargent, Impressionism, Chinese Scrolls, and Abstract Expressionism) showed marked differences between the two samples. Collectively, this suggests the scale may be most appropriate for use in artistically engaged samples. Our results also identify avenues for future development of the scale, including expanding the pool of terms used and revisiting the number of response options.

Keywords: art; aesthetics; aesthetic fluency scale; assessment; art education; item response theory

Measuring Art Knowledge:

Item Response Theory Analysis of the Aesthetic Fluency Scale

People's knowledge of the arts uniquely shapes how they understand and interact with art (Specker et al., in press). The Aesthetic Fluency Scale (Smith & Smith, 2006), one of the most prominent scales for measuring art knowledge, was developed to measure people's art knowledge via assessing self-reported understanding and knowledge of artists and art styles, periods, and concepts. For each of 10 items, people judge where—from never knowing the artist or art idea existed to possessing detailed knowledge of the artist or art idea—their expertise lies.

In their initial administration of the scale with visitors to the Metropolitan Museum of Art, Smith and Smith (2006) found the items ranged in the amount of knowledge visitors reported and that higher aesthetic fluency total scale scores were associated with increased age, number of museum visits, and art history training. Since its introduction, the Aesthetic Fluency Scale has been used to examine people's knowledge of visual art. Much of this work has used this scale with museum visitors and art experts (e.g., Mullennix & Robinet, 2018; Smith & Smith, 2006, Specker et al., in press), but researchers are increasingly using this measure to assess art knowledge in undergraduate student samples (e.g., Atari et al., 2020; Cotter et al., 2017; Donato, 2019; Fayn et al., 2018; Silvia, 2007, 2013; Silvia & Barona, 2009).

In the present research, we use an Item Response Theory (IRT) approach to the Aesthetic Fluency Scale to better understand its psychometric properties in people with and without art experience. An IRT approach treats scale items as the unit of analysis and permits the examination of specific qualities of the items, such as how hard each is or how well each distinguishes between people with differing levels of the construct measured by the scale. Because the Aesthetic Fluency Scale was developed and initially tested with museum visitors but

is being used more widely with undergraduate students not studying art, understanding the properties of this scale within different groups of participants is important. Here we examine the properties of the scale's items in a large sample and explore potential avenues for future scale refinement.

Method

Participants

Participants were 3,233 people who completed the Aesthetic Fluency Scale. The data were compiled from multiple projects involving undergraduate students with a variety of non-arts majors, art and art history students, and art museum visitors (see Table 1 for descriptions of the individual samples and demographic information). This sample was divided into two groups—undergraduate students with majors unrelated to the visual arts (n = 1,791) and artistically engaged individuals (i.e., students studying visual art, art museum visitors; n = 1,442)[1]. The purpose of using these subsamples was to examine the properties of the scale within the population the scale was developed for use with (i.e., artistically engaged individuals) and the properties of the scale in an increasingly studied population (i.e., undergraduate students with majors unrelated to the visual arts).

Aesthetic Fluency Scale

All participants completed the Aesthetic Fluency Scale (Smith & Smith, 2006). This scale contains 10 items referring to different artists (i.e., Mary Cassatt, Isamu Noguchi, John Singer Sargent, Alessandro Botticelli, Gian Lorenzo Bernini), artistic styles (i.e., Impressionism,

[1]Museum visitors and students studying visual art were grouped together because there were not enough art students to examine them as a separate group. The arts students were grouped with the museum visitors (rather than with the non-arts students) due to their engagement in visual art-related activities.

Fauvism, Abstract Expressionism), or art concepts (i.e., Egyptian Funerary Stelae, Chinese Scrolls). Participants indicated their familiarity with each of the items on a 5-point Likert scale (0 = "I have never heard of this artist or term"; 1 = "I have heard of this but don't really know anything about it"; 2 = "I have a vague idea of what this is"; 3 = "I understand this artist or idea when it is discussed"; 4 = "I can talk intelligently about this artist or idea in art"). The scale score is calculated by summing responses to all items, so the possible scale scores range from 0 to 40.

Graded Response Model

We used a graded response model to estimate parameters. The graded response model (GRM; Samejima, 1969) handles data with rankings or ordered polytomous categories such as responses to attitudinal assessments. From this model, we can obtain two types of estimates for each item: *discrimination parameters* and *boundary location parameters*. The discrimination parameter (a) indicates how well items differentiate between people with higher versus lower aesthetic fluency ability—higher values indicate greater discrimination ability. This can be seen in item category response functions (ICRFs) as lines with steeper slopes. Each item has its own discrimination parameter. The item boundary location parameters indicate the ability level (θ)—here, the level of aesthetic fluency—at which the probability of indicating a higher level of ability is 0.50[2]. The number of item boundary parameters estimated depends on the number of response options. For the Aesthetic Fluency Scale, there are four item boundary location parameters (b_1-b_4) estimated for each item because there are five response options. For example, b_1 represents the ability level at which the probability of responding with a 1 (i.e., "I have heard of this but don't really know anything about it") or greater is equal to 0.50, b_2 represents the

[2] Note that the item boundary location parameters do not indicate the ability level at which providing a specific response is equal to a probability of 0.50 as is the case for dichotomous IRT models.

ability level at which the probability of responding with a 2 or greater is equal to 0.50, and so forth.

As θ is a latent variable in IRT models, it should be assigned a scale. Typically, the θ-scale is set by assuming that the group is a sample from a population that is normally distributed with a mean of 0 and a standard deviation of 1. Therefore, a θ of 0 represents average ability and a θ of 1, for example, indicates being one standard deviation above average.

Additionally, we can obtain the *information* for each item and the complete scale in the form of item or test information curves. Item and test information are used to determine how much information each item and the entire scale provide for the estimation of ability along the entire scale. The peak of these curves indicates the ability level for which the item is most informative. For the present project, information tells us how much information each item and the scale as a whole provide for the estimation of aesthetic fluency levels.

Linking Parameter Estimates

Linking is a statistical procedure that adjusts for differences in group ability by placing the parameter estimates from one group onto the scale of another group. As previously mentioned, when estimating parameters, we assume our sample comes from a population that is normally distributed with a mean of 0 and standard deviation of 1 (i.e., the θ-scale). Because we estimated two IRT models—one for the artistically engaged individuals and one for the non-arts students—each set of parameter estimates is placed upon its own θ-scale with a mean of 0 and standard deviation of 1; however, it is conceivable that the groups could differ in ability levels (i.e., the θ-scale means for two groups are non-equivalent), making comparisons between groups challenging. Linking adjusts for these differences by putting estimates on a common scale and

allows for comparison between groups. Therefore, the item parameter estimates from the non-arts student sample was placed onto the scale of the artistically engaged sample.

To complete the linking process, we used the Stocking and Lord (1983) test characteristic curve method, which compares the difference between each test characteristic curve (TCC) on two θ scales (i.e., non-arts students versus artistically engaged) and returns linking constants *A* and *B* for use in completing the linear scale transformation to link the two groups. The *A* constant reflects the slope of the transformation whereas the *B* constant reflects the intercept of the transformation. To obtain the new parameter estimates for the non-arts student sample on the artistically engaged sample scale, their separately calibrated old parameters were transformed using the following equations:

$$a_{New} = \frac{a_{Old}}{A}$$

$$b_{New} = A \times b_{Old} + B$$

If the groups are equivalent in their level of aesthetic fluency, the linking constants would be $A = 1$ and $B = 0$. If the scale operates similarly within the two subsamples, the linked item parameter estimates for the non-arts students and artistically engaged individuals should be similar to one another; if there are substantial differences between the estimates, this suggests evidence of differential item functioning, meaning that there are factors related to group membership that influence people's probability of providing a particular response (AERA et al., 2014).

Results

Because the GRM is a unidimensional IRT model, we first performed principal components analyses for the two samples to confirm that the Aesthetic Fluency Scale measures a single dimension. Analyses were completed using the *prcomp* function in the *stats* package in R (R Core Team, 2020). The scree plots (see Figure 1) for each sample suggest that the majority of

the variance can be accounted for with one dimension; however, if using a cutoff criterion of eigenvalues greater than one, analyses for the non-arts student sample indicates retaining two components and for the artistically engaged sample indicate retaining four components. Coefficient omega, an index of internal consistency appropriate for use with ordinal scales (Peters, 2014), was acceptable for the artistically engaged (Ω = .90) and non-arts student subsamples (Ω = .76). Collectively, these indices suggest proceeding with unidimensional analyses is appropriate, which is in line with similar analyses conducted by Smith and Smith (2006) in their initial investigation of the scale.

A GRM was used to calibrate the discrimination and boundary location parameters for the 10 items. All calibrations were completed using the *mirt* package (Chalmers, 2012) in R. Descriptive statistics for all items are available in Table 2. Parameter estimates for the artistically engaged individuals in Table 3 and for non-arts students in Table 4—estimates reported for the non-arts students have been linked to the artistically engaged sample scale. Figures 2-5 display the ICRFs (Figures 2-3) for the artistically engaged and non-arts student samples, respectively, and item (Figure 4) and test information (Figure 5) for the artistically engaged sample.

Artistically Engaged Sample

Participants in the artistically engaged sample had an average aesthetic fluency score of 15.92 (*SD* = 9.16, *range* = 0 – 40). People indicated the greatest amount of knowledge of terms related to art styles and similar amounts of knowledge for the artist and art concept terms. Specifically, people endorsed the highest knowledge of Impressionism (*M* = 2.87, *SD* = 1.04) and Abstract Expressionism (*M* = 2.19, *SD* = 1.24) and the least knowledge of Isamu Noguchi (*M* = .94, *SD* = 1.30).

Item discrimination parameters ranged from 0.97 to 2.60—the least discriminating item was Chinese Scrolls (a = 0.97) and the most discriminating item was Mary Cassatt (a = 2.60; see Table 3). Collectively, artist items were the most discriminating item type and art concept items were the least discriminating item type.

Within the artistically engaged sample, the items demonstrated a range of difficulty (see Table 4 and Figure 2). For all items (except Isamu Noguchi), the first boundary parameters (b_1) were negative, indicating that people with below average aesthetic fluency were able to report some level of knowledge of the terms (i.e., respond with a 1 or higher). Of the ten items, Impressionism (b_1 = -2.25) and Abstract Expressionism (b_1 = -1.65) required the lowest aesthetic fluency level to endorse knowledge of the term. Indicating greater than a slight familiarity with a term (i.e., reporting having at least a vague idea of what the term is) required average or slightly above average ability for most items, however. Endorsing a high level of knowledge (i.e., reporting the ability to talk intelligently about the term) generally required an aesthetic fluency level over one standard deviation above average. Impressionism was the only term with a final boundary less than 1 (b_4 = .59); the two art concept terms—Egyptian Funerary Stelae (b_4 = 2.97) and Chinese Scrolls (b_4 = 3.89)—were the only items to have parameters greater than 2, suggesting someone would need to be nearly 3 standard deviations above average in their aesthetic fluency to indicate a high level of familiarity with these terms.

The test information (see Figure 5) suggests the aesthetic fluency scale provides the most information for people near average aesthetic fluency and little information for those at extremely low (i.e., $\theta < -2$) or high (i.e., $\theta > 2$) levels of aesthetic fluency. But the individual items showed variability in information provided (see Figure 4). The Egyptian Funerary Stelae and Chinese Scrolls items provided minimal information at all levels of aesthetic fluency. The

artist items were most informative for people with near average aesthetic fluency, with Mary Cassatt being the most informative artist item. For the art style items, Impressionism and Abstract Expressionism provided the most information for people with low aesthetic fluency and Fauvism provided the most information for people with slightly below average aesthetic fluency.

Non-Arts Student Sample

Non-arts students had an average aesthetic fluency score equal to 7.13 (*SD* = 5.42, *range* = 0 – 40), which is substantially and significantly less than the artistically engaged participants ($t(3,231) = 50.23$, $p < .001$, $d = 1.68$, 95% CI: [1.62, 1.73]). Like the artistically engaged sample, people indicated greater knowledge of art style terms and less knowledge of artist and art concept terms. Students indicated the most knowledge regarding Impressionism (*M* = 1.77, *SD* = 1.12) and Abstract Expressionism (*M* = 1.54, *SD* = 1.16) and the least amount of knowledge regarding Isamu Noguchi (*M* = .10, *SD* = .42). In fact, the median response for all but three terms (Impressionism, Chinese Scrolls, and Abstract Expressionism) was 0, meaning most students reported not having heard of most of the terms, unlike the artistically engaged individuals who only had a median response of 0 for one item (Isamu Noguchi).

The item parameter estimates for the non-arts students were linked to the artistically engaged sample using the *plink* package (Weeks, 2010) in R. The linking constants for the non-arts student sample were *A* = 0.66 and *B* = -1.11, meaning that the non-arts student sample had a mean ability of -1.11 and standard deviation of 0.66 relative to the artistically engaged sample with a mean ability of 0 and standard deviation of 1. The linked estimates (see Table 4) are on the artistically engaged sample θ scale, meaning that these estimates have been adjusted for differences in group ability. After linking, the linked parameter estimates for the non-arts

students should be close to those for the artistically engaged sample[3]. Substantial differences in these estimates would suggest that there are factors due to group membership—not aesthetic fluency level—that differentially influence how members of the artistically engaged and non-arts student samples respond to those specific items.

Most items had linked parameter estimates similar to those obtained for the artistically engaged sample; however, there were notable differences in parameter estimates for four items—John Singer Sargent, Impressionism, Chinese Scrolls, and Abstract Expressionism (see Table 4 and Figure 3). On average, non-arts students reported having less knowledge about three of these items than artistically engaged individuals; both samples indicated similar levels of knowledge of Chinese Scrolls (see Table 2). The John Singer Sargent item was more discriminating and required lower levels of aesthetic fluency to report high levels of familiarity (indicated by b_2, b_3, and b_4) in the artistically engaged sample. The Impressionism item was more discriminating in the non-arts student sample, but the samples did not differ in their item boundary parameter estimates for this item. The Chinese Scrolls item was more discriminating in the non-arts student sample, and non-arts students required lower levels of aesthetic fluency than the artistically engaged individuals to indicate higher levels of familiarity. Similarly, the Abstract Expressionism item was also more discriminating in the non-arts student sample, and at the highest level of familiarity (i.e., b_4), the non-arts students required lower levels of aesthetic fluency to indicate an ability to talk intelligently about the term that did the artistically engaged.

Discussion

[3] Because item and test information are calculated using the item parameter estimates, information for the non-arts student sample is also expected to be similar to estimates provided for the artistically engaged sample.

The present research used an IRT approach to examine the psychometric properties of the Aesthetic Fluency Scale (Smith & Smith, 2006) in samples of artistically engaged individuals and non-arts students. The results suggest that the items vary in their difficulty, with items related to art styles being easiest and art concept and artist items being more difficult, and how well they discriminate between people with low and high levels of aesthetic fluency. The scale appears to be best equipped for people who have aesthetic fluency levels near the artistically engaged sample's mean.

Unsurprisingly, the artistically engaged individuals had higher aesthetic fluency scores than did the non-arts students. For all but one item—Chinese Scrolls (see Table 2)—the artistically engaged sample reported greater familiarity than the non-arts student sample; for Chinese Scrolls, the two samples reported similar levels of familiarity. This pattern is largely to be expected—a component of most art education programs involves art history coursework, increasing the likelihood of exposure to the terms, and museumgoers may engage in informal learning through exposure to different artworks, artists, and museum programs during their visits (Smith & Smith, 2006).

These analyses have several implications for future use of the Aesthetic Fluency Scale. First, there are implications for which populations the scale is appropriate. There was a large difference between the two samples examined here—the artistically engaged indicated greater aesthetic fluency overall and greater familiarity with all but one of the items than the non-arts students. The non-arts students appeared to struggle with the difficulty of the scale—for seven of the ten items, the median response was never having heard of the term, the lowest possible option. Six percent of the non-arts students reported never hearing of any of the terms, and over 75% of the sample had an overall score less than 10 out of 40; only 2% of the artistically

engaged sample reported never hearing of any terms and 29% of the artistically engaged had an overall score less than 10.

This pattern suggests that the Aesthetic Fluency Scale in its current form suffers from a floor effect with non-arts students, and researchers interested in measuring aesthetic fluency and art knowledge may consider using a different scale in this population (e.g., the Vienna Art Interest and Art Knowledge scale; Specker et al., in press). This also presents the opportunity for additional development of the Aesthetic Fluency Scale to include additional art style items (as these were the easiest item type) or names of artists more widely known (e.g., Monet, Picasso) to better measure aesthetic fluency in non-arts students. Alternatively, researchers who are using this scale in its current form with non-arts students may consider log-transforming people's scores to reduce the influence of the students who do report high levels of aesthetic fluency and account for the positively skewed distribution (Bland et al., 2013).

Additionally, four items (John Singer Sargent, Impressionism, Chinese Scrolls, and Abstract Expressionism) showed signs of differential item functioning, suggesting that factors other than someone's aesthetic fluency level differentially influences how a non-arts student and artistically engaged individual would respond to these items. This poses a threat to the validity of the interpretation and use of scores for non-arts student samples as factors other than aesthetic fluency level influence people's scores (AERA et al., 2014). Since this scale was developed for and initially tested in a museum visitor population (Smith & Smith, 2006), these differences between the artistically engaged and non-arts student samples may also suggest discontinuing use of this scale in its current form in non-arts student samples.

Second, IRT analyses allow us to evaluate individual items. One method of examining items is looking at information each item provides about someone's aesthetic fluency levels (see

Figure 4). Different items will often be more informative at different levels of aesthetic fluency (i.e., some items will tell us more about people with near average aesthetic fluency whereas others may be more informative for people with somewhat above average aesthetic fluency). In this study, however, the Egyptian Funerary Stelae and Chinese Scrolls items provided virtually no information at all levels of aesthetic fluency. Because a person's response to these items does not provide substantial information about their aesthetic fluency level, removing these items may improve the performance of the scale (e.g., Lalor et al., 2016). In the present samples, omission of these items does not markedly change the scale's reliability ($\Omega = .90$ for the artistically engaged sample and $\Omega = .73$ for the non-arts student sample).

Finally, IRT analyses allow us to examine the response options of the scale using the ICRFs (see Figures 2 and 3). The trace lines in these figures show the probability of giving a particular response to an item at a particular level of aesthetic fluency. In cases where the trace lines completely overlap, this suggests that one response is never the most likely response at any given aesthetic fluency level. Across the full continuum of aesthetic fluency levels, it appears that only three responses consistently have trace lines not fully overlapping with other lines (i.e., have aesthetic fluency levels when that response is the most likely). This may suggest that this scale could benefit from reducing the number of response options, perhaps from five to three, to more effectively assess aesthetic fluency (e.g., Muraki, 1993).

The Aesthetic Fluency Scale has been widely used to measure people's art expertise. Overall, our findings suggest this scale may be most appropriate for artistically engaged individuals, such as art museum visitors or students studying visual art. The brevity of the Aesthetic Fluency Scale makes it a museum-friendly tool for visitors with limited time (and patience) for research participation, but we often have fewer time-constraints with student

samples, allowing us to use more appropriate, albeit somewhat longer, measures of art expertise. Our findings also provide avenues for additional development of the Aesthetic Fluency Scale that may make it more student-friendly and to improve the precision of its measurement.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Atari, M., Afhami, R., Mohammadi-Zarghan, S. (2020). Exploring aesthetic fluency: The roles of personality, nature relatedness, and art activities. *Psychology of Aesthetics, Creativity, and the Arts, 14*(1), 125-131.

Bland, J. M., Altman, D. G., & Rohlf, F. J. (2013). In defence of logarithmic transformations. *Statistics in Medicine, 32,* 3766-3769.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29.

Cotter, K. N., Silvia, P. J., Bertamini, M., Palumbo, L., & Vartanian, O. (2017). Curve appeal: Exploring individual differences in preference for curved versus angular objects. *i-Perception,* March-April, 1-17.

Donato, F. V. (2019). *Visual ambiguity priming promotes uniqueness in art-viewing responses* [Unpublished doctoral dissertation]. Texas Tech University.

Fayn, K., Silvia, P. J., Erbas, Y., Tiliopoulos, N., & Kuppens, P. (2018). Nuanced aesthetic emotions: Emotion differentiation is related to knowledge of the arts and curiosity. *Cognition and Emotion, 32*(3), 593-599.

Lalor, J. P., Wu, H., & Yu, H. (2016). Building an evaluation scale using item response theory. In *Proceedings of the Conference on Empirical Methods in Natural Language*

Processing. Conference on Empirical Methods in Natural Language Processing (Vol. 2016, p. 648). NIH Public Access.

Mullenix, J. W. & Robinet, J. (2018). Art expertise and the processing of titled abstract art. *Perception, 47*(4), 359-378.

Muraki, E. (1993). Information functions of the generalized partial credit model. *ETS Research Report Series, 1993*(1), i-12.

Peters, G. -J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist, 16*(2), 56-69.

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, 17.

Silvia, P. J. (2007). Knowledge-based assessment of expertise in the arts: Exploring aesthetic fluency. *Psychology of Aesthetics, Creativity, and the Arts, 1*(4), 247-249.

Silvia, P. J. (2013). Interested experts, confused novices: Art expertise and the knowledge emotions. *Empirical Studies of the Arts, 31*(1), 107-115.

Silvia, P. J. & Barona, C. M. (2009). Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empirical Studies of the Arts, 27*(1), 25-42.

Smith, L. F. & Smith, J. K. (2006). The nature and growth of aesthetic fluency. In P. Locher, C. Martindale, & L. Dorfman (Eds.), *New directions in aesthetics, creativity, and the arts* (pp. 47-58). Amityville, NY: Baywood.

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (in press). The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts.*

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*(2), 201-210.

Weeks, J. P. (2010). plink: An R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software, 35*(12), 1-33.

Table 1

Demographic Information for the Samples

	Collection Description	***N***			**Gender**	**Age**
		Total	**Artistically Engaged**	**Non-Arts Student**		
Sample 1	Participants were recruited as part of a course research participation option at a mid-sized Southeastern U.S. university	1,365	48	1,317	26.28% Male, 73.21% Female, 0.15% Unreported	*M* = 19.19 *SD* = 2.39 *Range* = 17 – 49
Sample 2	Participants were recruited during their visit to an art museum in New York City	960	960	0	31.20% Male, 55.22% Female, 13.58% Unreported	Under 18 = 42 18 – 24 = 96 25 – 34 = 124 35 – 44 = 125 45 – 54 = 158 55 – 64 = 169 65 – 74 = 135 75+ = 59
Sample 3	Participants were recruited for being psychology or art history students at a large Austrian university or for their expertise in art history	620	146	474	27.69% Male, 63.47% Female, 8.53% Unreported	*M* = 23.85 *SD* = 9.20 *Range* = 17 – 81
Sample 4	Participants were recruited during their visit to an art museum in the Southeastern U.S.	159	159	0	38.51% Male, 55.90% Female, 5.59% Unreported	*M* = 31.50 *SD* = 19.60 *Range* = 18 – 88
Sample 5	Participants were recruited during their visit to an art museum in the Netherlands	129	129	0	46.92% Male, 53.07% Female	*M* = 47.00 *SD* = 17.06 *Range* = 17 – 80

Non-Arts Student Sample	Includes undergraduate students who are not studying the visual arts	1,791	0	1,791	28.59% Male, 70.80% Female, 0.61% Unreported	*M* = 20.01 *SD* = 4.07 *Range* = 17 – 81
Artistically Engaged Sample	Includes art museum visitors and students studying the visual arts	1,442	1,442	0	32.04% Male, 58.74% Female, 9.92% Unreported	*M* = 43.62 *SD* = 18.54 *Range* = 17 – 88

Note. Because Sample 2 collected age information in age ranges, the midpoint for each range category was used in calculating the Artistically Engaged Sample age statistics.

Table 2

Descriptive statistics for Aesthetic Fluency Scale responses

	Full Sample	Artistically Engaged	Non-Arts Students
	M (*SD*), *Median*	*M* (*SD*), *Median*	*M* (*SD*), *Median*
Mary Cassatt	.78 (1.30) Med. = .00	1.55 (1.52) Med. = 1.00	.19 (.62) Med. = .00
Isamu Noguchi	.46 (1.00) Med. = .00	.94 (1.30) Med. = .00	.10 (.42) Med. = .00
John Singer Sargent	1.06 (1.29) Med. = 1.00	1.80 (1.46) Med. = 2.00	.50 (.75) Med. = .00
Alessandro Botticelli	1.17 (1.31) Med. = 1.00	1.90 (1.35) Med. = 2.00	.60 (.94) Med. = .00
Gian Lorenzo Bernini	.83 (1.21) Med. = .00	1.45 (1.37) Med. = 1.00	.37 (.79) Med. = .00
Fauvism	.81 (1.18) Med. = .00	1.34 (1.38) Med. = 1.00	.41 (.79) Med. = .00
Egyptian Funerary Stelae	.85 (1.12) Med. = .00	1.36 (1.23) Med. = 1.00	.46 (.83) Med. = .00
Impressionism	2.26 (1.21) Med. = 2.00	2.87 (1.04) Med. = 3.00	1.77 (1.12) Med. = 2.00
Chinese Scrolls	1.23 (1.09) Med. = 1.00	1.24 (1.16) Med. = 1.00	1.21 (1.03) Med. = 1.00
Abstract Expressionism	1.82 (1.24) Med. = 2.00	2.19 (1.24) Med. = 2.00	1.54 (1.16) Med. = 1.00
Total Scale Score	11.05 (8.53) Med. = 8.00	15.92 (9.16) Med. = 15.00	7.13 (5.42) Med. = 6.00

Table 3

Discrimination and boundary location parameter estimates for artistically-engaged individuals

	Discrimination (*a*)	Boundary 1 (b_1)	Boundary 2 (b_2)	Boundary 3 (b_3)	Boundary 4 (b_4)
Mary Cassatt	2.60	-.31	.00	.44	1.24
Isamu Noguchi	2.12	.18	.66	1.18	1.87
John Singer Sargent	2.46	-.73	-.18	.31	1.20
Alessandro Botticelli	2.49	-.86	-.40	.29	1.41
Gian Lorenzo Bernini	1.89	-.47	.12	.75	1.81
Fauvism	2.10	-.32	.23	.79	1.73
Egyptian Funerary Stelae	1.13	-.79	.25	1.42	2.97
Impressionism	2.23	-2.25	-1.58	-.66	.59
Chinese Scrolls	.97	-.79	.51	1.92	3.89
Abstract Expressionism	1.69	-1.65	-.80	.09	1.42

Table 4

Discrimination and boundary location parameter estimates for non-arts students

	Discrimination (*a*)	Boundary 1 (b_1)	Boundary 2 (b_2)	Boundary 3 (b_3)	Boundary 4 (b_4)
Mary Cassatt	2.69	-.02	.44	.81	1.39
Isamu Noguchi	2.50	.34	.89	1.58	1.76
John Singer Sargent	.85	-.55	1.96	3.25	4.55
Alessandro Botticelli	1.86	-.78	-.01	.61	1.72
Gian Lorenzo Bernini	2.26	-.37	.20	.70	1.63
Fauvism	2.47	-.54	.07	.71	1.44
Egyptian Funerary Stelae	1.49	-.45	.40	1.36	2.39
Impressionism	4.89	-1.87	-1.29	-.68	.06
Chinese Scrolls	2.15	-1.66	-.83	.08	1.29
Abstract Expressionism	4.16	-1.67	-1.11	-.54	.17

Note. Estimates reported have been linked to the artistically engaged sample.

Figure 1. Principal components analysis scree plots

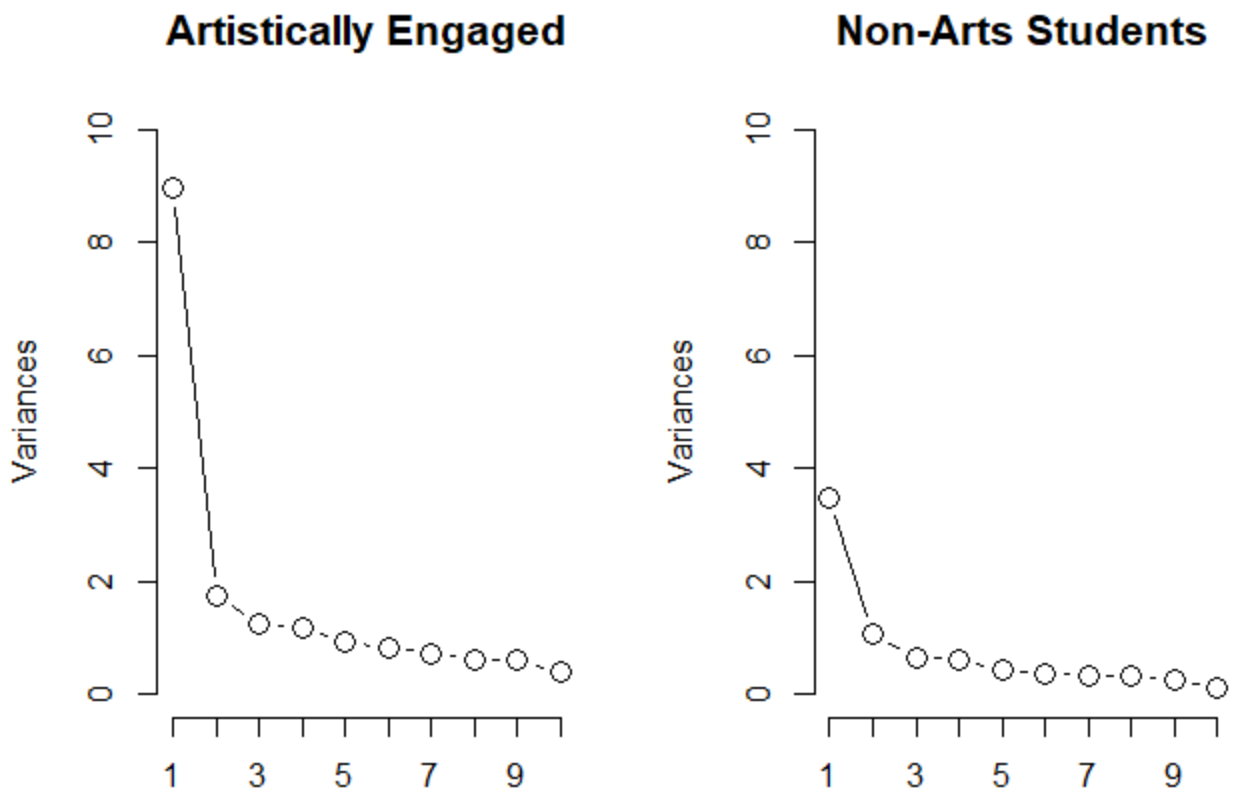

Figure 2. Item category response functions for the artistically engaged sample

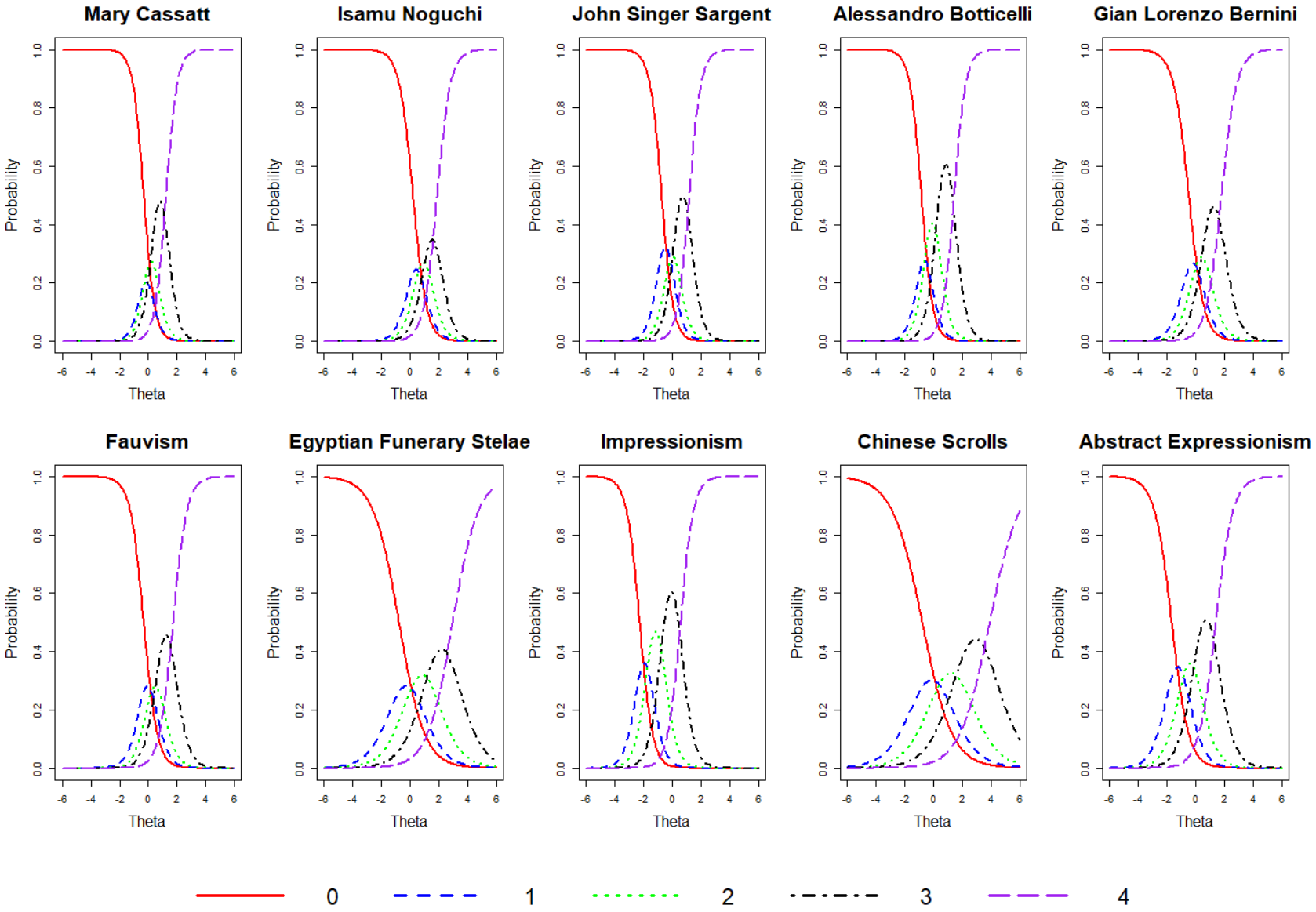

Figure 3. Item category response functions for the non-arts student sample

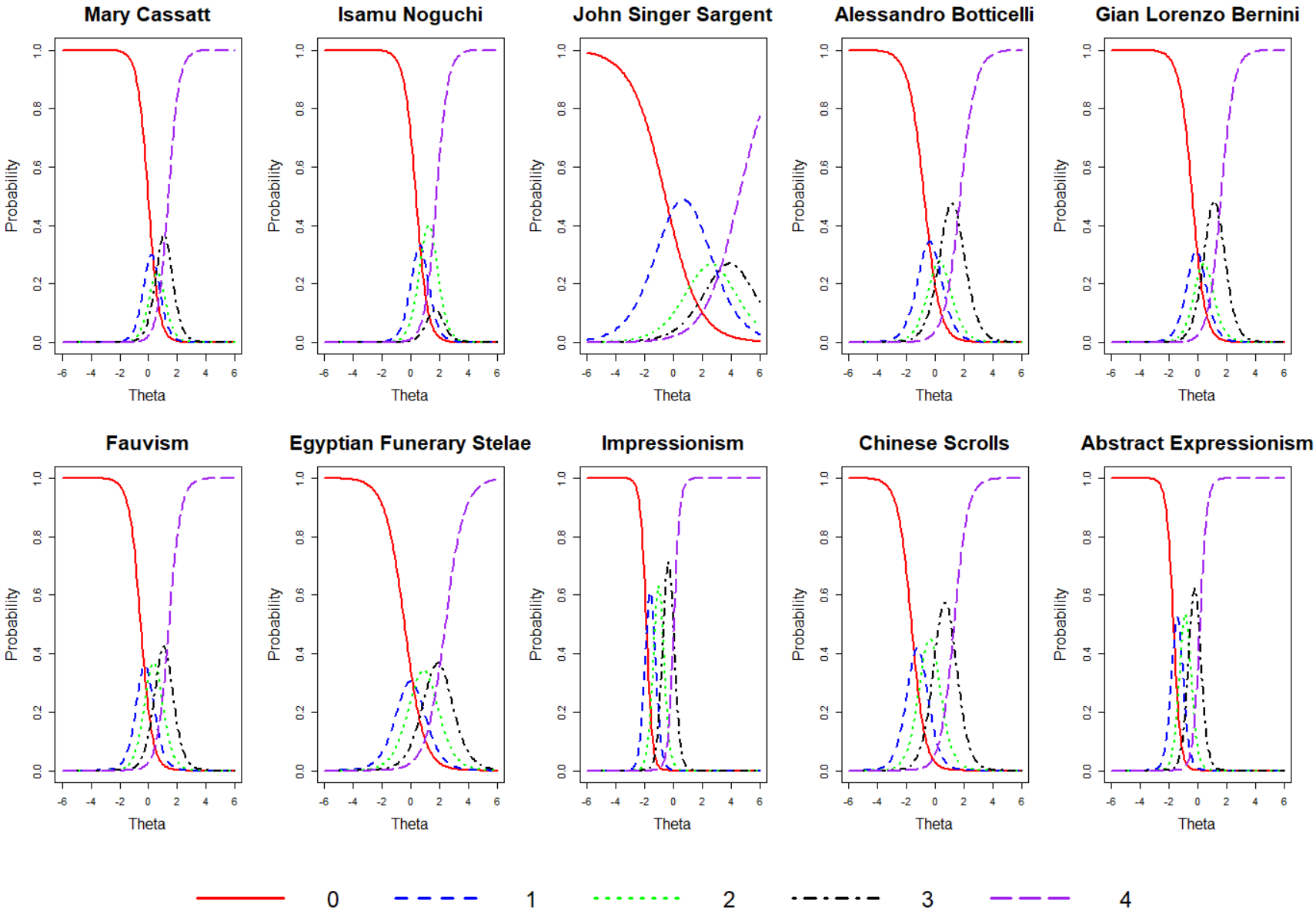

Figure 4. Item information for artistically engaged sample

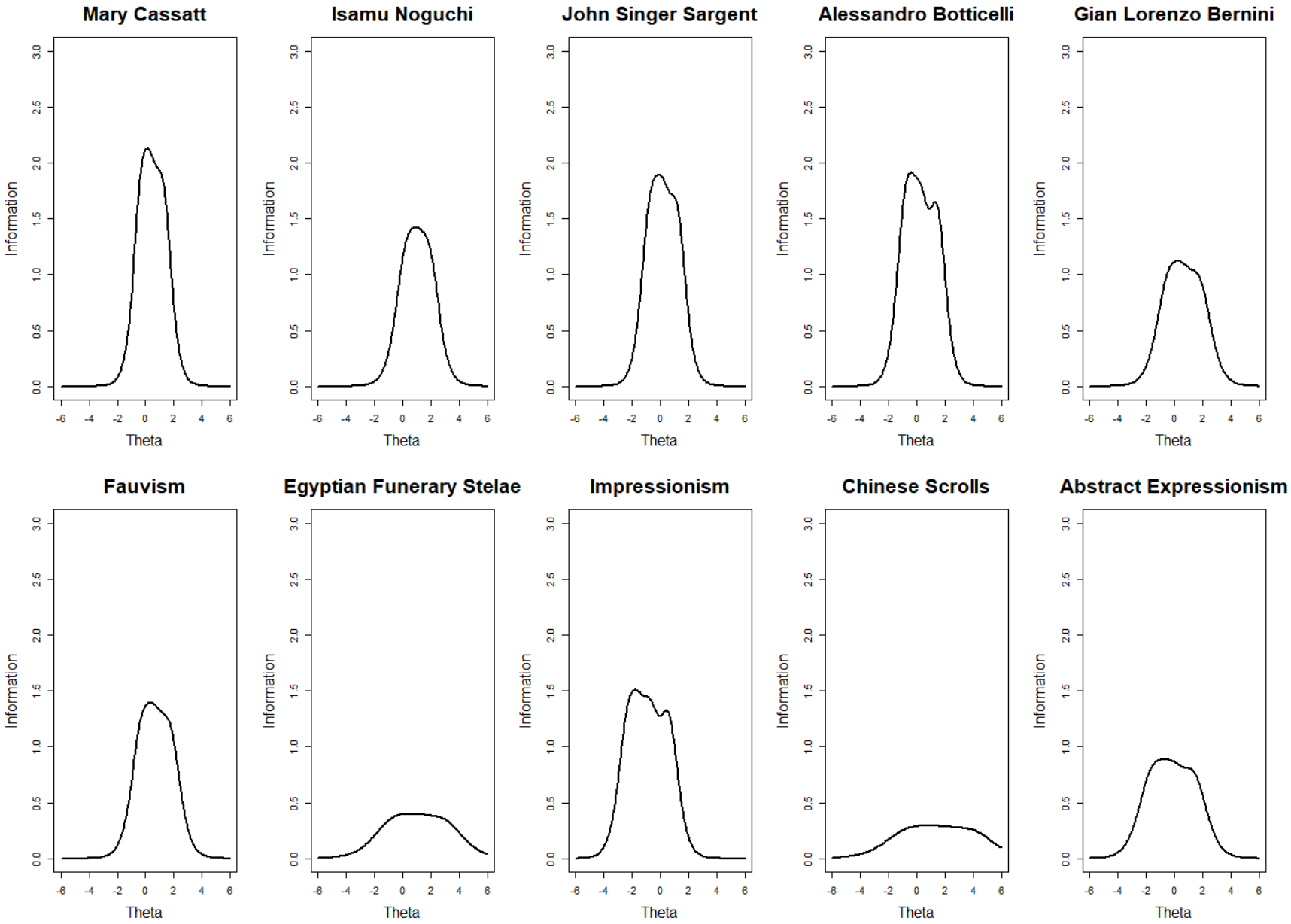

Figure 5. Test information for the artistically engaged sample