

Rule-based Aptitude Measurement: Artistic Judgment

Nikolaus Bezruczko, Ph.D. and Ambra Borgognoni Vimercati

Measurement and Evaluation Consulting

Introduction

In the 1920s educators and psychologists became aware of artistic judgment's importance for improving vocational selection and career development. They emphasized testing artistic judgment to select persons best suited for professions and occupations ranging from dentistry and architecture to window display and hair dressing (McAdory, 1929). (But what about plastic surgery, interior decorating, and film making? The instrumental influence seems almost limitless!) Unfortunately, valid and effective tests proved elusive for most of the twentieth century.

In retrospect, a major limitation was artistically dismal visual images serving as test items. Poor items in fact lead to some horrible disasters. The Meier Art Tests (1940), for example, presented old-world, masterpiece reproductions with an explicitly adulterated counterpart. Persons with high artistic judgment presumably would select the masterpieces. The artistic community (artists, museums, galleries, and so on) reacted with such shock and disbelief at this obvious violation of artistic dignity that they mocked and discredited all further attempts to test artistic judgment. To avoid embarrassment, commercial publishers abandoned artistic judgment testing, which to this day is hardly emphasized despite its obvious social benefits. The few contemporary tests purporting to measure artistic judgment have inconsistent, vague, or unknown reliability and validity.

Empirical Challenges

Twentieth century social researchers did substantially advance in understanding how visual images appeal to preference. Birkhoff, a noted mathematician, proposed that two image characteristics, complexity and uniformity, are functionally related to visual preference (1932). He systematically manipulated polygon sides to illustrate these characteristics but left empirical verification to others. Later, Hans Eysenck (1940, 1941) picked up this line by factor analyzing visual preferences for polygons. He found "T" and "K", described T as a general Taste factor, and K as a bipolar factor that distinguishes between artists and nonartists. Unfortunately, Eysenck's contributions were only based on raw score correlations and weak true score methods which ultimately would undermine his effort to test artistic judgment aptitude. Because visual artists logically appear at the extreme high end of the artistic judgment aptitude distribution where raw score distortions are greatest, a test derived from Eysenck's T, Visual Aesthetic Sensitivity Test (VAST) (Götz, 1981), was constantly plagued with validity and reliability problems. [Linear group differences in distribution tails only need to be 50% as large as their raw score differences to be statistically significant (Wright & Masters, 1982; see also Wright, 1999).]

In the 1960s, Berlyne experimentally showed that image complexity follows a curvilinear preference function (1971, 1974). Visual preference and image complexity monotonically increase until complexity reaches a maximum, then preference steadily declines. Unfortunately, an outcome of his research was incredible confusion concerning complexity's influence on artistic

judgment. Because his studies did not include artists, Berlyne and most other social researchers incorrectly assumed that preference for complexity is indicative of higher artistic judgment. Contemporary social researchers are typically astounded to learn that Berlyne's complexity function is *inversely* related to artistic judgment. In fact, complexity has profoundly different effects on artists and nonartists. Not grounded on objective measuring methods, this confusion concerning complexity and artistic judgment continues in contemporary research.

In addition to using "bad" art, twentieth century artistic judgment research suffered from:

- sample dependent, norm-based statistics
- nonlinear raw scores
- vaguely defined constructs
- deficient validity studies

Contemporary Advances

In the early 1980s, Johnson O'Connor Research Foundation (JOCRF), the oldest and largest aptitude testing organization in the United States, received a special gift from Christian A. Johnson Endeavor Foundation to develop new tests for its aptitude battery. Sensitive to artistic judgment's influence on occupations, JOCRF undertook to develop a reliable and valid artistic judgment test. JOCRF recruited the first author to research prior efforts, then to design an objective model for testing artistic judgment. Over the course of eight years, the outcome of this effort was remarkably successful. Exploiting the discriminant implications of Eysenck's K factor, he adapted a stochastic sampling model proposed by Attneave (1959) to construct rule-based images that varied only in complexity and redundancy. Moreover, these images emulated the contemporary art style known as Minimalism. According to Eysenck's K factor, dichotomously scoring image preferences that systematically vary in complexity should distinguish between artists and nonartists. Thirty-five items were bound into booklets, the Visual Design Test (VDT; 1987), which JOCRF extensively studied for psychometric properties in their testing offices (1989, 1990). VDT was finally validated with professional artists and other studies examined developmental implications among school children, and differences between artists and nonartists (Bezruczko & Schroeder, 1989, 1990, 1991, 1994, 1995; Schroeder & Bezruczko, 1990, 1998).

VDT Figurative Project

Despite this extraordinary success testing artistic judgment, JOCRF was reluctant to include VDT in its standard test battery. The extreme minimalism of VDT images was never accepted by the artistically staid foundation, and its staff frequently commented that VDT Abstract images "just don't look like art." In 1999 JOCRF asked Ambra Borgognoni Vimercati, an Italian fresco artist in Rome, to paint figurative images derived from rule-based VDT Abstract images. The purpose was to establish whether figurative paintings conforming to traditional, Western art standards and derived from an algorithm could be psychometrically reliable and valid. She painted 20 canvases in five styles: Renaissance, Baroque, Impressionism, Fauvism, and Surrealism, manipulating only their complexity. (In each style, she painted four variations of a theme.) VDT Figurative images, however, fundamentally differ from VDT Abstract because she defined artistic judgment as "coherent integration of multiple image elements in a theme" and systematically introduced unnecessary elements to increase complexity. Complexity in VDT Abstract is

only manipulated by element frequency. This difference in operational definition motivates the present study.

Using the same item format as VDT Abstract, less-complex figurative paintings were paired with more-complex paintings and published as VDT Figurative (Bezruczko, Borgognoni Vimercati, & Calipari, 2000). JOCRF then administered these items in their testing offices in Chicago, New York, Boston, and Dallas to study their measurement properties. Figure 1 (see p. 28) shows VDT Abstract, Figurative, and their co-calibration to a common scale using WINSTEPS (Linacre & Wright, 2000). In this article, we present preliminary results from this study.

The sample was predominantly white, educated, and socioeconomically upper middle class. Their arts background, however, was generally modest and none were professionally-trained artists. Co-calibration shows these dichotomously scored items are well targeted on 244 JOCRF examinees with generally acceptable fit values. [Infit and outfit meansquares are 1.00 and .99, respectively, for both persons and items. Person separation = 2.35 (reliability = .85) and item separation = 4.32 (reliability = .95).] (An issue concerning poor item fit for several hard items was accommodated by removing four persons with implausible preferences. We speculate high item difficulty elicited some random responses.) Residual analysis shows a systematic factor structure accounting for 12 percent of residual variation. (Factor 1 explains 5.04 of 42 residual variance units.) Residual variation revealed a second factor we call Uniformity that is consistent with prior factor analyses.

Figure 2 (see p. 29) shows that this variable is defined by a coherent order of figurative items. Impressionism, Surrealism, and Baroque styles define the lower portion while Renaissance and Fauvism define the upper portion. Because items are scored for their agreement with artist preference, the lower portion shows items on which it was *easy* to agree with artists, while items higher on this variable are much harder. Consequently, persons high on this continuum, though artistically untrained, tend to agree with preferences of professional artists, suggesting they have a natural “eye” for art. Approximately 12 percent fell in this category. The clustering of styles on this continuum reveals that Surrealism and Impressionism are substantially easier to pass (agree with professional artists), while Fauvism and Renaissance styles tend to be hardest.

Insights and Conclusions

This research departs from traditional artistic judgment studies because it examines visual preference within the quantitative rigor of an invariant linear framework. There it applies knowledge acquired from studies of synthetic, rule-based images to paint figurative art. A surprising result was that the influence of complexity on agreement with artists does not dramatically differ between abstract and figurative art. Co-calibration of VDT Abstract and Figurative items show them intermingled and supported by reasonable fit values.

Complexity, however, does appear to profoundly influence preference for representative and nonrepresentative figurative art. Complexity’s influence on preference for representational figurative styles such as Renaissance, Baroque, and Fauvism is predictable – more complex is always preferred over less complex. However, when this nonartist sample was presented nonrepresentational figurative styles, such as Impressionism and Surrealism, they inexplicably abandoned more-complex images and preferred less-complex images! We speculate that items using nonrepresentational images “overload” viewers with uninterpretable content, leading to rejection of more-complex alternatives, and inadvertently increasing preference for less-complex, art-preferred images. Representational, figurative images, on the other hand, provide familiar, meaning-

ful content, which allows complexity to stimulate viewers to higher preference but less agreement with professional artists.

These preliminary results are being extended by analyses that compare person measures for VDT Abstract and Figurative items separately to assess construct comparability (Bezruczko, in press). Further validity studies will verify consistency with professional artist preferences. Our implementation of an invariant measurement structure will make cross-cultural generality of these results relatively easy to examine. This structure will also be useful for assessing applicability of the underlying complexity principle, that is, association between complexity and nonartist preferences across other art styles. This capacity to replicate offers rare opportunities to extend knowledge about artistic judgment and better understand its role in professions and occupations. Finally, the linearity and invariance of this framework will be important to understanding the functional relationship between complexity and representationalism.

References

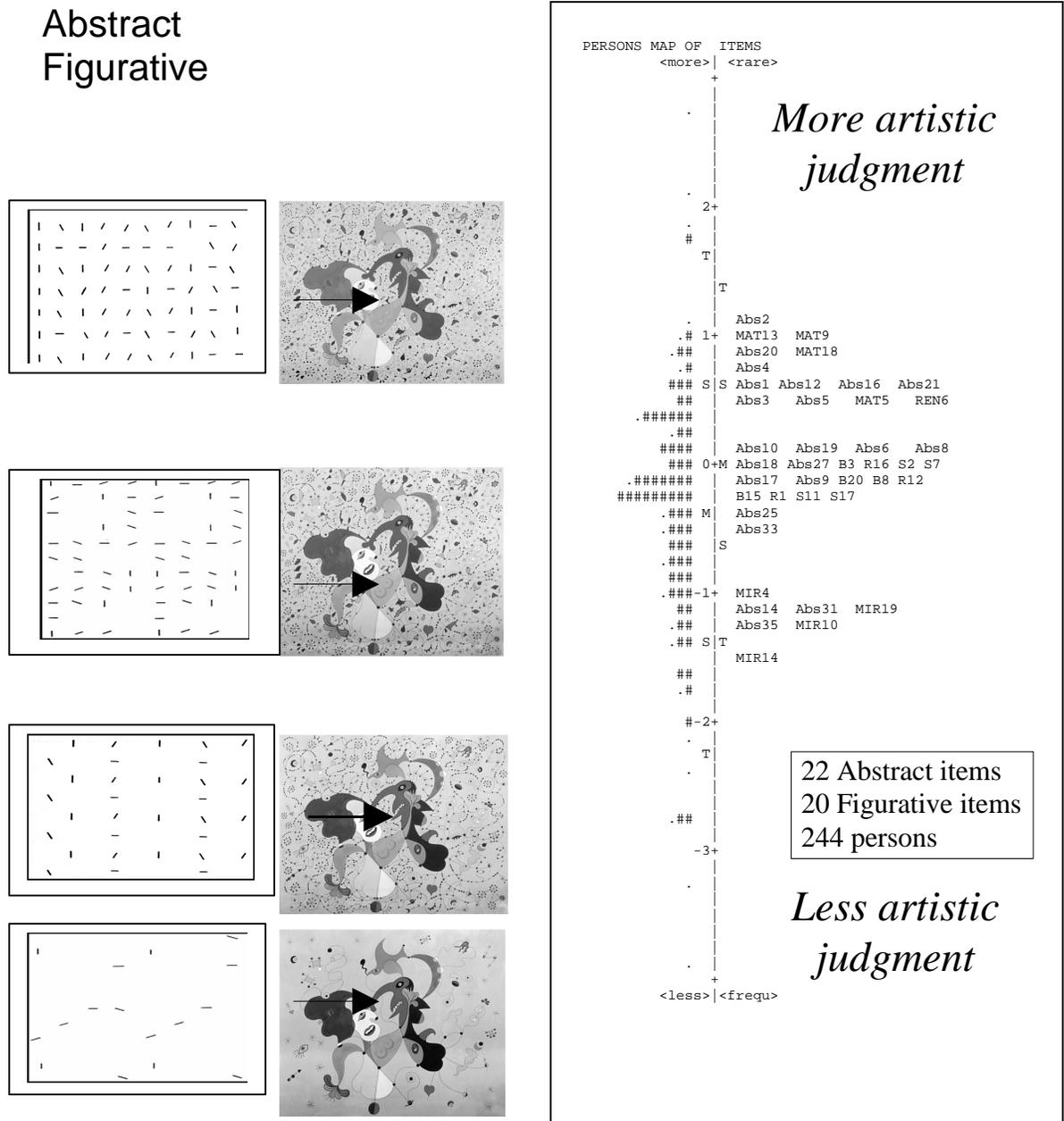
- Attneave, F. (1959). Stochastic composition processes. *Journal of Aesthetics and Art Criticism*, 17, 503-510.
- Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.
- Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Washington, DC: Hemisphere Publishing Corporation.
- Bezruczko, N., Borgognoni Vimercati, A., & Callipari, P. (2000). *Visual Designs Test: Figurative pilot edition*. Rome, Italy: Courage Press.
- Bezruczko, N., & Schroeder, D. H. (1987). *Visual designs test: Abstract*. Chicago: Johnson O'Connor Research Foundation, Inc.
- Bezruczko, N., & Schroeder, D. H. (1989). *Technical report 1989-2. Artistic Judgment Project I: Internal-structure analyses*. Chicago: Johnson O'Connor Research Foundation, Inc.
- Bezruczko, N., & Schroeder, D. H. (1990). *Technical Report 1990-1. Artistic Judgment Project II: Construct validation*. Chicago: Johnson O'Connor Research Foundation, Inc.
- Bezruczko, N., & Schroeder, D. H. (1991). *Technical Report 1991-1. Artistic Judgment Project III: Artist validation*. Chicago: Johnson O'Connor Research Foundation, Inc.
- Bezruczko, N., & Schroeder, D. H. (1994). Differences in visual preferences and cognitive aptitudes of professional artists and nonartists. *Empirical Studies of the Arts*, 12, 19-39.
- Bezruczko, N., & Schroeder, D. H. (1995). Development of visual preferences in art-trained and non-art-trained schoolchildren. *Genetic, Social, and General Psychology Monographs*, 122, 179-196.
- Birkhoff, G. P. (1932). *Aesthetic measure*. Cambridge, MA: Harvard University Press.
- Eysenck, H. J. (1940). The general factor in aesthetic judgments. *British Journal of Psychology*, 3, 94-102.
- Eysenck, H. J. (1941). Type factors in aesthetic judgments. *British Journal of Psychology*, 31, 262-270.
- Götz, K. O. (1981). *Visual aesthetic sensitivity test*. Dusseldorf: Concept Verlag.
- Linacre, J. M. (2000). *WINSTEPS: Rasch analysis for all two-facet models, version 3.04*. Chicago: MESA Press.
- McAdory, M. (1929). *The construction and validation of an art test*. New York: Bureau of Publications, Columbia University.
- Meier, N. C. (1940). *Meier Art Tests: Part I. Art judgment*. Iowa City: Bureau of Educational Research and Service, University of Iowa.

- Schroeder, D. H., & Bezruczko, N. (1990). *Visual preference dimensions that distinguish artists and nonartists*. Paper presented at Annual Convention of American Psychological Society, Dallas.
- Schroeder, D. H., & Bezruczko, N. (1998). *Technical report 1998-1. Artistic judgment: Review of project and study of school children*. Chicago: Johnson O'Connor Research Foundation, Inc.
- Wright, B. D. (1999). Fundamental measurement in psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement*. Mahwah, NJ: Lawrence Erlbaum.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Notes

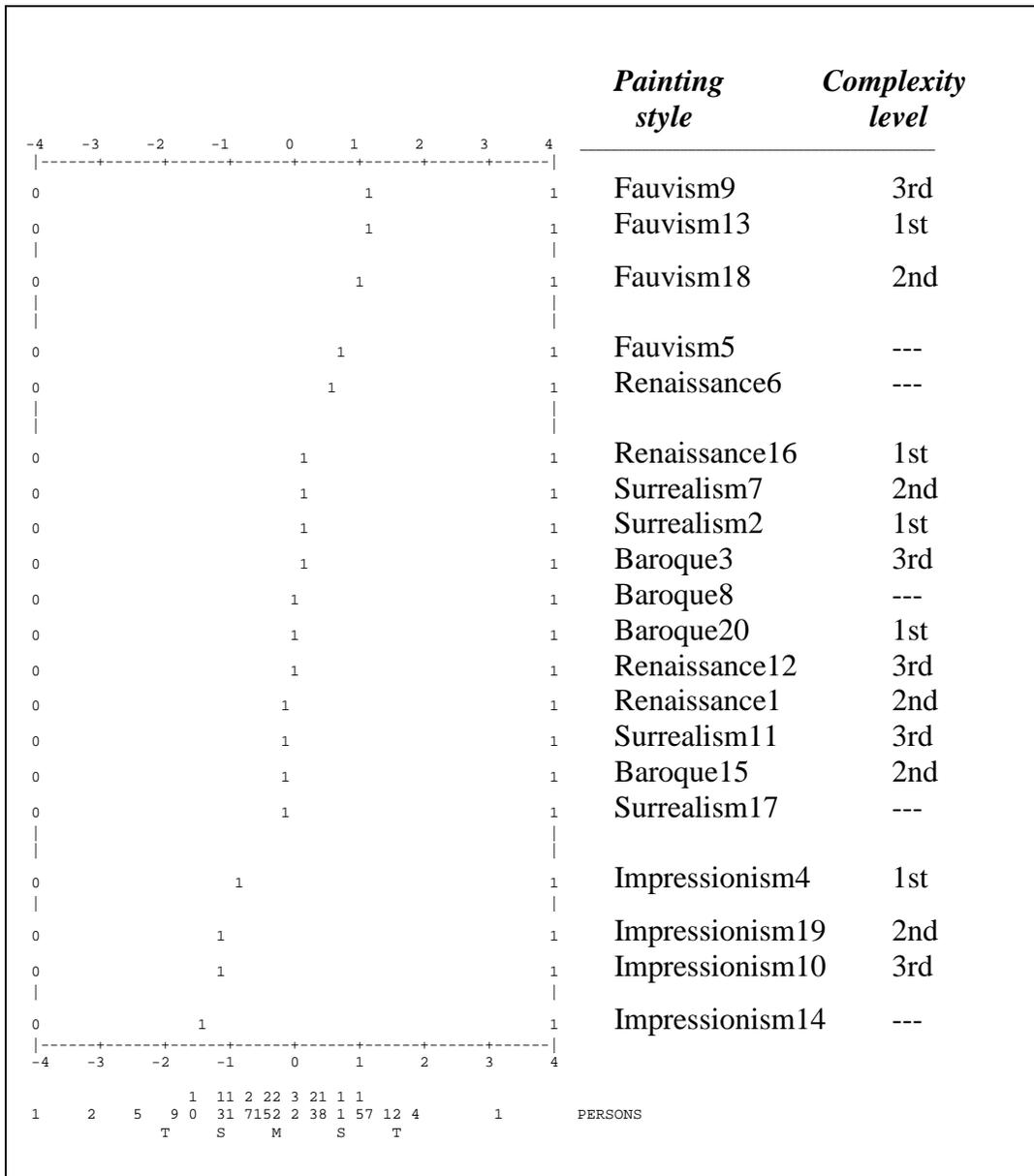
1. Nikolaus Bezruczko, Ph.D., is a MESA graduate and co-founder of the Chicago Objective Measurement Table (COMET).
2. Ambra Borgognoni Vimercati is a professional artist specializing in fresco restoration based in Rome, Italy. Her current work emphasizes emic characteristics in visual art.
3. We gratefully acknowledge cooperation of Johnson O'Connor Research Foundation, and in particular Dr. David H. Schroeder, Research Manager, and Robert F. Kyle, Director of Research, in collecting data for this research. We are also indebted to Professor Ben Wright for permitting us to collect responses to mockups for VDT Figurative paintings from his students. Their responses were instrumental in producing effective figurative items. Sample VDT Abstract and Figurative items are at: www.artisticjudgement.com.

Figure 1. Co-calibration of VDT Abstract and Figurative



Note. Abstract images are constructed by an algorithm that randomly assigns elements to a design. Painted figurative images conform to complexity level in corresponding VDT Abstract. Both VDT Abstract and Figurative images were co-calibrated to define a common artistic judgment aptitude construct.

Figure 2. Figurative item profiles



Note. All forced-choice items are keyed by artist preference. Nonrepresentational figurative styles define lower portion of this variable, indicating items on which it is easier to agree with artists. Upper portion is defined by representational figurative images that are very difficult. We speculate they identify persons commonly referred to as “natural artists.”