A Meta-Analysis of the Relationship Between Intelligence and Visual “Taste” Measures

Nils Myszkowski  
Pace University

Pinar Çelik and Martin Storme  
Université Paris Descartes - Sorbonne Paris Cité

What makes individuals experts in judging aesthetic value is actively researched in a variety of ways. In the visual domain, one classical paradigm—used in “T” (for Taste) tests (Eysenck, 1983)—consists in comparing one’s evaluative judgments of beauty with a standard judgment—provided by consensual or expert agreement. The association between general intelligence (g) and performance in “T” tests has been investigated since over 70 years (Eysenck, 1940; Myszkowski, Storme, Zenasni, & Lubart, 2014), but has led to a variety of results, from negative weak to positive strong correlations. We aimed at clearing the resulting confusion through a meta-analysis of the correlations observed in the literature (k = 23, N = 1,531). We found a significant positive weak to moderate correlation between g and “T” (r = 0.30, 95% CI [confidence interval] = [0.23, 0.36], z = 9.00, p < .001), suggesting that common cognitive processes are involved in both g and “T”. Reinforcing this conclusion, no publication bias was found through the regression test, and none of the tested moderators—year of publication, gender, age, “T” measure, and g measure—had a significant effect on the correlation.

**Keywords:** aesthetic sensitivity, taste, intelligence, aesthetic ability, meta-analysis

Despite the difficulty of finding objectivity in aesthetic judgments, empirical aesthetics, and psychology researchers have relentlessly attempted to identify the features of consensually preferred stimuli (e.g., Locher & Nodine, 1989; Marković & Gnozdenovi, 2001; Reber, 2002; Silvia & Barona, 2009; Vitz, 1966). While a lot of this research has been focused on aesthetic universals (e.g., Arnheim, 1954; Graves, 1951) and their flexibility (e.g., McManus, 2005), the person perspective (Jacobsen, 2006) is interested in individual differences in aesthetic judgments (e.g., Barron, 1953; Eysenck & Furnham, 1993; Furnham & Avison, 1997; Furnham & Bunyan, 1988; Rawlings, Twomey, Burns, & Morris, 1998). A special—and controversial (Eysenck, 1997; Gear, 1986)—approach to these individual differences consists in studying the extent to which some individuals may be better equipped to judge aesthetic stimuli.

**What Is “T”?!**

Although there are different approaches to aesthetic ability—including aesthetic expertise (Chatterjee, Widick, Sternschein, Smith, & Bromberger, 2010; Plucker, Kaufman, Temple, & Qian, 2009; Silvia, 2007; Smith & Smith, 2006), sensitivity to complexity (Barron & Welsh, 1952), aesthetic chills (Silvia & Nusbaum, 2011), and exploration tendencies (Nodine, Locher, & Krupinski, 1993)—one that has recently regained interest (Chamorro-Premuzic & Furnham, 2004; Furnham & Chamorro-Premuzic, 2004; Myszkowski, Storme, & Zenasni, 2016; Myszkowski et al., 2014; Summerfeldt, Gilbert, & Reynolds, 2015) is the aesthetic sensitivity—or “T”—approach. This approach consists in measuring and studying the ability to identify the aesthetic value of stimuli by directly presenting stimuli that vary in aesthetic quality, such quality being either measured through consensual or expert agreement (Child, 1962; Eysenck, 1940), or manipulated through the “controlled alteration” (Meier, 1928, p. 188) of stimuli (that consists in altering the stimuli to make it of lesser aesthetic quality). Despite the psychometric challenges that such an approach raises, notably in terms of content validity (Myszkowski et al., 2014), researchers have recently found interest in “T”’s suggested relations with various dispositional variables, including personality traits (Chamorro-Premuzic & Furnham, 2004; Freo & Eysenck, 1995; Myszkowski et al., 2014), creativity (Myszkowski et al., 2014), artistic training (Frie & Eysenck, 1995), obsessive-compulsive disorder tendencies (Summerfeldt et al., 2015), and intelligence (Bezruczko & Freo, 2011; Myszkowski et al., 2014).

Although “T” previously existed (Meier, 1928), evidence supporting the relevance of the “T” factor approach was found in the early 1940s, when, using exploratory factor analyses of the judgments of a variety of aesthetic stimuli, a general factor was found to explain most of the variability in individual visual aesthetic preferences across wide arrays of stimuli (Eysenck, 1940, 1941a, 1941b, 1942, 1968). Originally named “T” for good Taste (Eysenck, 1940, 1941a, 1941b, 1942, 1968).
This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of individual users and is not to be disseminated broadly.

senck, 1983), this “T” factor translates Eysenck’s findings—and later Child’s (1962)—that individuals differ in their general degree of agreement with consensus aesthetic judgments of stimuli of various nature. In other words, people tend to agree with one another on what they consider of superior aesthetic value, and the individuals who agree more with such consensus preferences tend to do so across a wide variety of types of stimuli. Thus, people differ consistently across most visual domains in their degree of agreement with consensus judgments: That degree of agreement is “T.” Later, “T” started becoming referred to as “aesthetic sensitivity” (Child, 1962, 1964), and was defined by Child (1964, p. 49) as the “extent to which a person gives evidence of responding to relevant stimuli in some consistent and appropriate relation to the external standard.”

Since Meier’s early development of the controlled alteration method (Meier, 1928), many measures of aesthetic sensitivity have been developed with this method and studied: The Meier Art Tests (Meier, 1940, 1963), the Design Judgment Test (Graves, 1948, 1951), and the Visual Aesthetic Sensitivity Test (Götz, 1985; Götz, Borisy, Lynn, & Eysenck, 1979); all three are still used in current research (Chamorro-Premuzic & Furnham, 2004; Furnham & Chamorro-Premuzic, 2004; Myszkowski et al., 2014). Although alternate methods to measure “T” have been discussed, like Lifton’s (1961) suggestion to ask participants to indicate whether two paintings were of the same artist, or like the PJB Test (Bamossy, Scamoon, & Johnston, 1983), and although some of the initial “T” measures did not involve stimulus manipulation (Child, 1962; Eysenck, 1940), the measure of “T” through standardized testing is typically achieved through the “controlled alteration” method (Meier, 1928). In the case of these tests, pairs or triads of similar visual stimuli are created, one stimulus remaining unaltered, while the other(s) are altered to present defaults—support for content validity may be found in the consensus among a community sample, and/or in the agreement of experts (Götz, 1985). Study participants have to indicate which of the two or three presented designs is of the best quality (Götz et al., 1979; Myszkowski et al., 2014) or most appealing (Child, 1962).

“T” and g

At the time that “T” was conceptualized, the dimensionality of intelligence was at the heart of the debate for individual differences psychology researchers (Burt, 1940; Eysenck, 1939; Spearman, 1939; Thurstone, 1938), some of them—particularly Eysenck—being involved in research on both the general factor of intelligence, g, and “T” (Myszkowski et al., 2016). Therefore, it is not surprising that, since then, not only g and “T” have been investigated using similar statistical methods—both g and “T” strongly rely on exploratory factor analyses in their original conceptualization (Eysenck, 1941a, p. 91; Spearman, 1939)—but also possible connections between g and “T” have been considered from the beginning: In the early characterization of “T,” Eysenck described “T” as “merely the manifestation of “g” when aesthetic material is used” (Eysenck, 1940, p. 101), leading to early searches for correlations between g and “T.”

Common Processes

Eysenck’s description of good taste as a facet of intelligence is cavalier—“T” and g measures being after all very different in nature. Nevertheless, there are cognitive processes involved in psychometric intelligence that “T” tasks may tap into.

Attention shifting. A first example of such common processes could be the inhibition of nonrelevant features and the attention-shift to relevant ones. Indeed, when presented with an aesthetic stimulus during a “T” task, the capacity of individuals to shift attention to specific relevant features of the stimulus may be crucial to their performance. Typically, in the case of the “controlled alteration” method used for “T” tests (e.g., Götz, 1985), to better perform, individuals have to shift their attention, from the attributes that are equivalent in both stimuli to the features that were altered—the very locations of the altered stimuli’s flaws. This tentative explanation of attention-shifting as a process common to g and “T” is consistent with the finding that art-trained viewers are better able to spot differences between an original stimulus and its altered version (Nodine et al., 1993), consistent with the identification of attention-shifting as a key element in the perceptual advantage of artists (Kozbelt & Seeley, 2007), and consistent with the finding that inhibiting tasks that require attention-shifting (e.g., Stroop tasks) are correlated with g measures (Friedman et al., 2006).

Reflective processing. Related to the idea that inhibition may be an important common process between g and “T,” and in line with the idea that art expertise is characterized by the tendency to adopt an aesthetic stance that is “emotionally distanced” (Leder, Gerger, Briebier, & Schwarz, 2014, p. 1138) when judging aesthetic stimuli, the observed correlations between g and “T” could be partially due to the ability to inhibit a mode of aesthetic processing that is focused on direct emotional arousal and the protection of existing knowledge, to facilitate a different mode of processing, which, although partially relying on emotions, would be more analytic and focused on knowledge acquisition. In this view, Cupchik (1995) contrasts the reactive model of aesthetic processing—where pleasure and emotional arousal have the most importance—with the reflective model—where emotional responses are elements that help in the analysis and interpretation of stimuli. This distinction can be related to the distinction made by Armstrong and Detweiler-Bedell (2008) between pursuing prevention goals (seeking to prevent confusion and maintain existing knowledge) and promotion goals (seeking to expand knowledge) when judging aesthetic stimuli. However, it should not be excluded that emotional factors play a role in “T” measures, and it should not be considered that higher “T” scores are achieved through the total inhibition of emotion: Processing complex aesthetic stimuli is not simply a problem-solving task (Muth, Heslinger, & Carbon, 2015), as it notably relies to a great extent on affective rewards associated to experiencing insights in judging stimuli. In this view, recent research relating personality traits to “T” measures (Myszkowski et al., 2014) notably indicated that openness to aesthetics, openness to fantasy, openness to feelings, and sensation seeking are related to higher “T” scores.

Goal management. Another potential process that could be commonly implied in g and “T” measures is goal management. Indeed, goal management—spawning subgoals from goals and tracking the success or not in pursuing each subgoal to achieve a higher goal—underlies success in g tests (Carpenter, Just, & Shell, 1990), and may also be implied in “T” type tasks. It can be argued that participants who are more able to spawn subgoals from a “T” test item (e.g., successively analyzing and evaluating symmetry,
organizational balance, line breaks, contrast, etc.) are better able to deconstruct the very process of evaluating the aesthetic quality of a stimulus into separate steps, thus focusing more distinctly on the different aspects, and achieving better performance in identifying flaws at each step. This could consequently lead to a higher ability to achieve the higher goal of identifying aesthetic quality, leading to higher performance in “T” tests. Once again, this should be nuanced by the nonequivalence between an aesthetic judgment task and a problem-solving task (Muth et al., 2015): It is possible that, when facing the aesthetic stimuli used in “T” tests, the ability to adopt a gestalt approach, rather than to decompose a problem into subproblems, is more efficient—especially when the “T” task is presented with a time limit. Alternatively, it is also possible that higher performance in “T” tasks is achieved when the subgoals that are pursued are not purely the result of a cognitive “management” process, but when instead individuals follow insights about what subgoals need to be addressed primarily.

**Abstract.** Finally, we could also propose abstraction—that consists of “the construction of representations that are only loosely tied to perceptual inputs and instead are more dependent on high-level interpretations of inputs that provide a generalization over space and time” (Carpenter et al., 1990, p. 428)—as a cognitive process that is both involved in psychometric intelligence tests and in “T.” Indeed, as participants analyze stimuli, they may need to detach themselves from direct perceptual inputs to use abstract rules and use the structural skeleton (Arneim, 1954) as a stronger basis for judgment. In this view, art experts are often found to be more able to focus on the elements of an aesthetic stimulus that are less directly available: They for example direct their attention more toward the relationships between objects than on the objects as individual elements (Nodine et al., 1993), and to focus on the order and dynamics of visual structure (Winston & Cupchik, 1992). In other words, individuals with higher “T” levels may use inputs of higher-order than individuals with lower levels, which would direct their attention on inputs that are more directly available in the stimulus.

**Observed correlations**

The question of a possible correlation between g and “T” has been investigated multiple times in the visual domain, on various samples (children and adults), and with various measures of both constructs (Bezruczko & Frois, 2011; Chamorro-Premuzic & Furnham, 2004; Child, 1962; Eysenck, 1940; Frois & Eysenck, 1995; Furnham & Chamorro-Premuzic, 2004; Götz, 1987; Götz et al., 1979; Myszkowski et al., 2014). However, it has not yet led to a clear answer: A range of negative or positive weak and nonsignificant correlations to strong positive correlations have been observed (Bezruczko & Frois, 2011; Child, 1962; Frois & Eysenck, 1995). These studies were conducted with relatively small sample sizes, with the largest sample size being 130 participants (Myszkowski et al., 2014). Consequently, the literature provides contradictory results about the relationship between g and “T” with approximately half of the studies yielding significant positive correlations and the other half yielding nonsignificant correlations.

Not only have the observed correlations and their significance varied, but the conclusions about the relationship between the two constructs have also varied, from discarding intelligence as an “irrelevant feature of the individual” (Eysenck, 1983, p. 230) for judgments made with “T” measures, to seeing in these correlations a sign that intelligence may facilitate aesthetic judgment through cognitive processes (Myszkowski et al., 2014). Facing mixed evidence for the direction and magnitude of the correlation between the two constructs, researchers have been—and are still—cautious in their conclusions, discussing psychometrical challenges (Bezruczko & Frois, 2011; Furnham & Chamorro-Premuzic, 2004; Iwawaki, Eysenck, & Götz, 1979; Myszkowski et al., 2014).

**The Present Research**

Our aim in this article was to provide, through a systematic literature review and meta-analysis, an estimate for the magnitude and direction of the correlation between the general factor of intelligence, g, and “T.” As we explained earlier, there is a lot of diversity in the research on the topic: Studies have been conducted (a) with various measures of g, (b) various measures of “T,” (c) on very different samples of children and adults from different countries, (d) have led to a wide range (from negative weak to positive strong) of correlations, which (e) has itself led to a wide array of conclusions.

As we have earlier explained that common cognitive processes may be involved in “T” and g measures, we hypothesized that “T” measures are positively correlated with intelligence measures. We believe that a demonstration through systematic review and meta-analysis can clear the confusion created by such heterogeneous research, provide guidelines for future research, as indicate that intelligence should not be discarded as playing a role in “T” measures.

**Method**

**Literature Search**

The PsycArticles, PsycInfo, and ERIC databases were searched independently by two researchers, using the following keywords, previously identified during a prereview of the literature: “(A)esthetic ability,” “(A)esthetic judgment,” “(A)esthetic sensitivity,” “Art judgment,” and “Art ability.” Furthermore, three additional searches were performed, with in each, as a keyword, one of the three main tests of visual “T”: the Visual Aesthetic Sensitivity Test (Götz, 1985), the Design Judgment Test (Graves, 1948), and the Meier Art Tests (Meier, 1940, 1963). In addition, all reference lists of the obtained articles were inspected for studies that could have been missing from the databases. The titles and abstracts were screened for the potential inclusion of a “T” and g measure. The search was not date restricted, but the publication date of the studies that met the inclusion criteria ranged between 1940 and 2014. The search was not country, age, culture, or language specific.

**Inclusion Criteria**

All the studies that were included in the meta-analysis had to meet the following criteria: (a) The studies included a judgment measure—either an ad hoc measure (e.g., Eysenck, 1940) or a standardized test (Götz, 1985; Graves, 1948; Meier, 1940, 1963)—of “T” that directly operationalized Child’s definition.
(Child, 1964, p. 49), which consists, as we earlier explained, of asking individuals to judge or compare stimuli based on their beauty or aesthetic pleasantness, and then confronting their judgments with external standards; (b) the stimuli that were used to measure “T” had to be visual stimuli of artistic or decorative nature (e.g., the participants were not asked to rate the beauty of individuals); (c) the research included an intelligence test of any sort; (d) reporting the relation between “T” and g, enough information (typically the correlation coefficient and sample size) was reported to estimate effect size and its variance; and (e) the participants were not sampled from a clinical population. The total number of included studies was 23, and the pooled total sample size was 1,531. The summary of the different measures used to measure “T” and g in the included studies is reported in Table 1.

### Data Input

We input the effect size and sample size in the R-packages 'metafor' (Viechtbauer, 2010; Viechtbauer & Viechtbauer, 2015) and 'meta' (Schwarzer, 2007; Schwarzer & Schwarzer, 2015). All but one of the included studies reported correlation coefficients and sample sizes, which were consequently directly input.

The study (Chamorro-Premuzic & Furnham, 2004) that did not report a g–“T” correlation coefficient investigated the relation between g and “T” through hierarchical regression (incrementally adding predictors to a regression model). In this study, the researchers entered intelligence as a predictor of “T” after having previously added art interests. They reported the incremental part of variance explained (\( R^2_{\text{Change}} \)) because of entering intelligence. We transformed the \( R^2_{\text{Change}} \) into a correlation coefficient metric by using its root square, before including it in the meta-analysis. Even though this study was not an outlier in the meta-analysis, it can be noted that the presence of art interests as predictors in the regression models of this study may have led to slightly misestimating the correlation between g and “T.”

Whenever the studies investigated and presented g–“T” correlations as different results obtained on different samples (Child, 1962; Frois & Eysenck, 1995; Götz, 1987; Götz et al., 1979), we included the results on each sample directly, without pooling them (because the samples were independent). In same-sample situations, however, we averaged correlation coefficients: When, in the same sample, the correlation between g and “T” was reported when “T” was both measured using the agreement with experts and with the agreement with consensus (Child, 1962), or when, in the same sample, a correlation between “T” and two different intelligence tests was presented (Bezruczko & Frois, 2011), we input the average of the two correlations and sample size.

### Statistical Analyses

The meta-analysis was performed using the ‘metafor’ function of the R-package ‘metafor’ (Viechtbauer, 2010; Viechtbauer & Viechtbauer, 2015) and with the ‘rma.uni’ function of ‘metafor’ (Viechtbauer & Viechtbauer, 2015). The reasons why two different packages were used were that they had different capabilities: Specifically, “meta” allowed more control over the plots presented in Figure 1 and Figure 2, but did not allow moderator analyses, which ‘metafor’ allowed. Besides, using two packages allowed the results of our statistical procedure to be cross validated.

**Model.** The heterogeneity in the research methods used—that we earlier underlined—clearly called for choosing a priori the random-effects estimation method of the correlation between g and “T.” For the random-effects model, we used the Hartung-Knapp-Sidik-Jonkman (Hartung, 1998; Hartung & Knapp, 2001a, 2001b; Sidik & Jonkman, 2005) estimation method, which has been showed to outperform the classic DerSimonian-Laird method (DerSimonian & Laird, 1986), especially in cases similar to ours where the number of studies is small (\( k = 23 \)) and there is heterogeneity (IntHout, Ioannidis, & Borm, 2014; Sidik & Jonkman, 2005).

**Heterogeneity.** Heterogeneity was measured through Higgins and Thompson’s \( I^2 \) (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003), which has been shown to be a more accurate and straightforward estimate of heterogeneity that is because of variation between studies than Cochran’s \( Q \) (Cochran, 1954).

### Table 1

**Summary of the Measures Used in the Included Studies**

<table>
<thead>
<tr>
<th>Publication</th>
<th>g measure</th>
<th>“T” measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eysenck (1940)</td>
<td>Verbal intelligence test (White, 1931)</td>
<td>Ad hoc measure</td>
</tr>
<tr>
<td>Child (1962)</td>
<td>“Scholastic Aptitude Test Verbal” (not referenced)</td>
<td>Visual Aesthetic Sensitivity Test (Götz et al., 1979)</td>
</tr>
<tr>
<td>Götz et al. (1979)</td>
<td>“IQ” (not further specified or referenced)</td>
<td>Visual Aesthetic Sensitivity Test (Götz, 1985)</td>
</tr>
</tbody>
</table>
Publication bias. Because the g–“T” relation is typically presented as a “side” result in many studies (e.g., Bezruczko & Frois, 2011; Child, 1962; Eysenck, 1940), there was no reason to expect that there could be a substantial publication bias. However, we investigated publication bias empirically through both the visual inspection of the contour-enhanced funnel plot (Peters, Sutton, Jones, Abrams, & Rushton, 2008) and the Egger regression test for publication bias (Sterne & Egger, 2005).

Additional moderation analyses. Although the main focus of this meta-analysis was the correlation between g and “T,” and although the reviewed literature did not indicate that the relation between g and “T” could be moderated by any factor, we decided to perform additional moderation analyses using ‘metafor,’ with the same estimation methods explained earlier. The reason for exploring for such moderation effects was to gather information on whether or not the correlation between the two constructs was generalizable to a variety of samples and measures. In other words, we wanted here to explore if there were signs that the heterogeneity of the observed correlations could be because of different characteristics of the samples and different methods, or if it could be imputable to sampling error.

We tested the moderation effect of (a) the year of publication, (b) the gender of the participants—some articles (Frois & Eysenck, 1995; Götz et al., 1979) reported correlations separately for males and females—(c) the mean age of the participants—that was not always available (Child, 1962; Eysenck, 1940), although when only the range was reported (Chamorro-Premuzic & Furnham, 2004; Furnham & Chamorro-Premuzic, 2004) we used the mean between the maximum and minimum—(d) the type of “T” measure—an ad hoc measure (Child, 1962; Eysenck, 1940) or a standardized test (Götz, 1985; Götz et al., 1979; Graves, 1948; Meier, 1940, 1963)—and (e) the type of g measure—verbal, nonverbal, or composite.

Results

g–“T” Correlation

We found a significant positive weak to moderate correlation between g and “T” (p = .30, 95% confidence interval (CI) [.23, .36], z = 9.00, p < .001).

The forest plot presented in Figure 1 gives an overview of the results of the meta-analysis. It presents each study through its effect size (the central point), the 95% CI of its effect size (the horizontal segment), and its weight in the meta-analytic random effects model (the size of the gray squares surrounding the central point). The fixed and random effects estimates of the correlation and their confidence intervals (respectively, the white and the black diamonds) are also plotted.

The weights of the studies did not indicate that a few studies contributed substantially more than the others, which could have biased the effect size estimate. However, the forest plot shows heterogeneity in effect size and statistical significance across the studies, with a number of confidence intervals including 0. The percentage of heterogeneity because of variation between studies was moderate (I² = 42.5%, 95% CI [5.7%, 65%]), which was expected, considering not only the inconsistency in the observed correlations that led to this meta-analysis, but also all the differences between these studies, in terms of procedures and sample characteristics.

The contour-enhanced funnel plot (reported in Figure 2) plots the size of each study—as recommended (Sterne & Egger, 2001), SE was used as a measure of study size—as a function of its observed effect size. It is inspected for asymmetry, potentially indicating a relation between study size and effect size, as well as for the potential suppression of nonsignificant findings. Here, the plot showed no asymmetry, as confirmed by the nonsignificance of the Egger regression test (t(21) = 1.20, p = .24), and did not suggest the suppression of nonsignificant findings. These results indicate no evidence of publication bias.

<table>
<thead>
<tr>
<th>Study and sample (sorted by year of publication)</th>
<th>Sample size</th>
<th>r</th>
<th>95% C.I.</th>
<th>Weight (Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eysenck (1940)</td>
<td>18</td>
<td>0.25</td>
<td>-.26 to .64</td>
<td>1.8%</td>
</tr>
<tr>
<td>Child (1962) - Standford</td>
<td>22</td>
<td>0.30</td>
<td>-.08 to .65</td>
<td>2.2%</td>
</tr>
<tr>
<td>Child (1962) - Yale</td>
<td>22</td>
<td>0.12</td>
<td>-.32 to .02</td>
<td>2.2%</td>
</tr>
<tr>
<td>Götz et al. (1979) - Boys</td>
<td>204</td>
<td>0.07</td>
<td>0.07 to .21</td>
<td>7.5%</td>
</tr>
<tr>
<td>Götz et al. (1979) - Girls</td>
<td>185</td>
<td>0.20</td>
<td>.06 to .34</td>
<td>7.0%</td>
</tr>
<tr>
<td>Götz &amp; Östergard (1987) - Private Gymnasium</td>
<td>102</td>
<td>0.37</td>
<td>.18 to .53</td>
<td>5.9%</td>
</tr>
<tr>
<td>Götz (1987) - Commercial College</td>
<td>73</td>
<td>0.62</td>
<td>.45 to .74</td>
<td>5.1%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Boys (10)</td>
<td>61</td>
<td>0.33</td>
<td>.09 to .54</td>
<td>4.6%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Boys (11)</td>
<td>49</td>
<td>0.28</td>
<td>.00 to .52</td>
<td>4.0%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Boys (12)</td>
<td>40</td>
<td>0.31</td>
<td>-.11 to .04</td>
<td>3.5%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Boys (13)</td>
<td>39</td>
<td>0.34</td>
<td>.02 to .68</td>
<td>3.4%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Boys (14)</td>
<td>36</td>
<td>0.21</td>
<td>-.13 to .50</td>
<td>3.2%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Boys (15)</td>
<td>40</td>
<td>0.13</td>
<td>-.19 to .42</td>
<td>3.5%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Girls (10)</td>
<td>50</td>
<td>0.26</td>
<td>-.02 to .50</td>
<td>4.1%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Girls (11)</td>
<td>49</td>
<td>0.29</td>
<td>.01 to .53</td>
<td>4.0%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Girls (12)</td>
<td>55</td>
<td>0.45</td>
<td>.21 to .65</td>
<td>4.3%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Girls (13)</td>
<td>51</td>
<td>0.07</td>
<td>-.21 to .34</td>
<td>4.1%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Girls (14)</td>
<td>50</td>
<td>0.25</td>
<td>-.03 to .49</td>
<td>4.1%</td>
</tr>
<tr>
<td>Frois &amp; Eysenck (1985) - Girls (15)</td>
<td>51</td>
<td>0.50</td>
<td>.26 to .68</td>
<td>4.1%</td>
</tr>
<tr>
<td>Furnham &amp; Chamorro-Premuzic (2004)</td>
<td>74</td>
<td>0.28</td>
<td>.06 to .48</td>
<td>5.1%</td>
</tr>
<tr>
<td>Chamorro-Premuzic &amp; Furnham (2004)</td>
<td>102</td>
<td>0.33</td>
<td>.14 to .49</td>
<td>5.9%</td>
</tr>
<tr>
<td>Beurczuklo &amp; Frois (2011)</td>
<td>48</td>
<td>0.50</td>
<td>.25 to .68</td>
<td>4.0%</td>
</tr>
<tr>
<td>Myshkowsli et al. (2014)</td>
<td>130</td>
<td>0.27</td>
<td>.10 to .42</td>
<td>6.0%</td>
</tr>
</tbody>
</table>

Fixed effect model: 1531 | 0.28 | [.23; .33] | 100%
Random effects model: 1531 | 0.30 | [.23; .36] | 100%

Figure 1. Forest plot with 95% confidence intervals and weights.
Moderators

None of the tested moderators—the year of publication, $F(1,21) = 0.98, p = .33$, the gender of the participants, $F(1,12) = 1.27, p = .28$, the age of the participants, $F(1,18) = 0.29, p = .60$, the type of “T” measure, $F(1,21) = 0.67, p = .67$, the type of $g$ measure, $F(2,20) = 1.27, p = .52$—had a significant effect on the correlation between the two constructs. Although these tests do not confirm null relationships (only the inability to reject null relationships), the obtained results suggest that the correlation between $g$ and “T” does not depend on any of these characteristics.

Discussion

For more than 70 years (Eysenck, 1940; Myszkowski et al., 2014), researchers have been interested in the correlation between intelligence and “T” scores in the visual domain. They have been largely confused by the variety—in terms of magnitude, statistical significance and even direction (e.g., Child, 1962)—of the correlations that have been observed between the two constructs.

We believe that the present results, obtained through a systematic review and meta-analysis, have cleared the confusion: Although about half of the studies found nonsignificant correlations, our effect size estimate allows to conclude that there is a positive correlation between $g$ and “T” measures, and that it is not of negligible magnitude (Cohen, 1988). More specifically, the estimated confidence intervals indicate that the correlation is small to moderate. We can add that it appears from our publication bias analyses that this result is probably not because of publication bias. Furthermore, no significant part of the heterogeneity in the observed correlations was explained by either year of publication, gender, age, the type of “T” test, or the type of $g$ test. This strengthens our point, as it suggests that the variability in the observed correlations is not due to the characteristics of the sample or the procedure: To the contrary, it suggests that such heterogeneity could be imputable to sampling error.

Implications

Although this is a meta-analytic investigation of correlational results, the most important implication of this study is that it suggests that common processes are implied in both psychometric intelligence and “T” measures. The present results do not indicate which processes these may be, but the fact that the type of $g$ measure (verbal, nonverbal, or composite) had no significant mod-
eration effect on the g–“T” correlation indicates that these processes are probably not specific to one single type of task.

Concerning these common g–“T” processes, when introducing the topic of this research, we advanced a (nonexhaustive) list of possibilities. Although it appears from the very instructions used (e.g., Götz, 1985) that “T” measures do not necessitate high abilities in maintaining and updating information in working memory—that are often described as the center of focus of psychometric intelligence tests (Belacchi, Carretti, & Cornoldi, 2010; Friedman et al., 2006)—psychometric intelligence tests actually involve other processes that “T” measures could tap into. More specifically, we proposed that “T” involves one or more of the following processes also involved in g (Carpenter et al., 1990; Friedman et al., 2006): Attention-shifting—high “T” individuals would be more able to shift their attention toward different elements, while low “T” individuals would remain more “stuck” on some elements that are not necessarily relevant—reflective processing—high “T” individuals would be able to adopt a mode of processing that aims at extending knowledge, while low “T” individuals would adopt a mode of processing that seeks direct emotional arousal—goal management—high “T” individuals would be more able to spawn subgoals from the task, being thus able to focus specifically on aesthetic features than low “T” individuals—and abstraction—high “T” individuals would be more focused on higher-order features, the structural skeleton, of the aesthetic stimuli, while low “T” individuals would focus more on the individual objects that are more directly available in the stimulus. Future research could investigate the individual cognitive processes involved in “T” tasks that could explain the correlation through our meta-analysis, using either a psychometrical approach—correlating individual differences in “T” with the performance at tasks that specifically measure the abilities associated with each cognitive process—or an experimental approach—using for example a dual-task paradigm, where individuals would have to perform two tasks simultaneously, one being a “T” type task, and the other task involving a specific cognitive process that would reduce the cognitive resources available for the “T” task.

We could add that, on one hand, recent research (Myszkowski et al., 2014) has indicated that, in addition to being correlated with g, “T” measures may be related to both openness and creativity, reinforcing the idea that artistic reception and creation use the same structural framework (Tinio, 2013), and that artists have higher perceptual skills (Kozbelt & Seeley, 2007; Kozbelt, Seidel, ElBassiony, Mark, & Owen, 2010); on the other hand, creativity, openness, and intelligence have been found to be also correlated (Feist, 1998; Gignac, Stough, & Loukomitis, 2004; Kim, 2008; Plucker, Esping, Kaufman, & Avita, 2015). The present evidence for a correlation between g and “T” therefore may be explained by potential presence of a common “intelligent” or “achievement” factor behind altogether g, “T”, creativity, and openness: We think that further investigation on how “T” relates to openness and creativity will enrich the understanding of how it relates to g.

Limitations

Our meta-analytic investigation has limitations. First, our investigation was only carried over studies that measured “T” in the visual domain with a specific procedure, which consists in comparing participants’ aesthetic judgments with standard judgments provided by consensual agreement or experts. This restricts our conclusion to that type of measure in particular, which is certainly quite specific and cannot fully encompass aesthetic ability, expertise, or fluency. As was recently noted (Myszkowski & Zenasni, 2016), the current state of research on these different constructs calls for a more multifactorial approach of aesthetic capacity—that would include not only “T” tests, but also the depth of aesthetic exploration tendencies (Koide, Kubo, Nishida, Shibata, & Ikeda, 2015; Nodine et al., 1993), art knowledge (Silvia, 2007; Smith & Smith, 2006), sensitivity to complexity (Barron & Welch, 1952), also named “K” (Eysenck, 1941b; Myszkowski et al., 2016), and the tendency to experience aesthetic chills (Silvia & Nusbaum, 2011)—that would probably altogether more accurately and comprehensively represent the ability to judge art, or “good taste,” than comparisons of individual taste with common or expert taste.

Another limitation is that “T” measures are known to face psychometric challenges (Liu, 1990), notably a lack of reliability (Bezruczko & Fris, 2011), which could have attenuated many of the correlations that were observed. Indeed, it may be that g and “T” are even more associated than we think, and future research should aim at improving “T” measures to gain a better accuracy in the estimation of the correlation between the two constructs.

Conclusion

The scatteredness of the literature on the relations between g and “T” called for the meta-analytic review provided here. Despite such variety, our review and meta-analysis have clearly indicated that there is a correlation that researchers should not overlook. Despite a very heterogeneous body of research, the main result is rather straightforward and leaves little doubt as to the presence of a correlation. However, we do not believe that researchers should consider the relation between g and “T” as completely established. To the contrary, the results of this meta-analysis raise new questions, and call for the further investigation of the cognitive processes involved in “T” measures.

References


INTELLIGENCE AND VISUAL TASTE: A META-ANALYSIS


INTELLIGENCE AND VISUAL TASTE: A META-ANALYSIS

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.

- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.

- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.

- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/authors/review-manuscript-ce-video.aspx.