Using Creativity to Reduce Ethnic Bias in College Admissions

James C. Kaufman
California State University at San Bernardino

College admissions testing typically focuses on grade point average (GPA) and SAT scores. Without disputing the importance of these predictors, one may wonder whether they are enough by themselves to determine eventual academic success. One possible additional construct, creativity, is examined via the lens of nonbiased assessment. It is argued that creativity can help reduce bias in two ways. First, adding creativity as a supplement to current assessments would present a truer reflection of a person’s overall intellectual abilities. Second, most studies have found that different ethnicities perform comparatively on creativity tasks, and some minority groups may be more likely to see themselves as creative. Including creativity as a component of standardized tests may, therefore, reduce stereotype threat.

Keywords: creativity, admissions, fairness, ethnicity, stereotype threat

When you say . . . there’s a test score gap, part of what I hear is that we have been taking pictures . . . using a particular kind of camera that seems to show white skin and middle class experience in all its glory. And that those with a different kind of experience don’t show up very well using this particular film, or this particular angle (Lani Guinier, as quoted in Chandler, 1999).

The tools used to determine college admissions should ideally reflect the purpose of higher education. Rigol (2003) highlights many different institutional philosophies about who should be admitted to college. Some colleges seek students who will succeed, a broad definition that could encompass students getting good grades, participating in campus life, or obtaining good jobs after college. Another perspective might be to desire students who are particularly good matches for the college. However, more recent approaches, such as those espoused by Bowen and Bok (1998), argue that it is equally important to create a diverse student body. One way to enhance diversity is to make sure that the information used for college admissions is as nonbiased as possible.

If an institution strongly values such diversity, are the current admissions tools the best ones to use? Currently, approximately 90% of undergraduate colleges, for example, require the SATs (or ACTs); these tests and high school GPA are the two most important determinants of college admissions (Zwick, 2007). The basic importance and quality of the SATs and GREs are not under much dispute; there are extensive studies that show validity for these instruments (Kuncel & Hezlett, 2007; Kuncel, Hezlett, & Ones, 2001; Sackett, Borneman, & Connelly, 2008). Instead, a larger question arises: are the SATs (or ACTs) are enough by themselves (or with GPA) to both determine eventual success in undergraduate or graduate programs and allow for a diverse student body? Even if the standardized tests are not specifically biased, they may not be the best way to attain nonbiased assessment. One way to attain this goal would be to add supplementary measures as a key part of the admissions process. One construct that stands out as a possible candidate is creativity.

Moving Toward Nonbiased Assessment

There are two ways that standardized tests are criticized for being potentially biased. A common layperson approach to criticizing tests as biased is to point to significant differences that occur between males and females and among ethnic groups on various tests of aptitude or ability. Indeed, a wide variety of measures of intelligence and ability have shown lower scores for African Americans and Hispanic Americans than for Caucasians (see Loehlin, 1999, and Weiss, Chen, Harris, Holdnack, & Saklofske, 2010, for overviews). Standardized tests, such as the SAT, ACT, GRE, and Advanced Placement (AP) exams, have shown similar patterns of discrepancy among ethnic groups (Camaras & Schmidt, 1999; Morgan & Maneeckshana, 1996). Some researchers argue that these measures reflect actual differences (e.g., Herrnstein & Murray, 1994; Jensen, 1998). Others point to the discrepancy between socioeconomic status and opportunities across ethnicities (Rogers, 1996; Sternberg, 1996), whereas still others argue that current ability measures do not incorporate enough aspects of intelligence to truly reflect a person’s “global” ability (Sternberg, Kaufman, & Grigorenko, 2008).

Researchers who advocate psychometric approaches to bias in testing take a more sophisticated view of the problem and do not accept the notion that just because two groups perform differently on a mental test then, therefore, the test itself must be in error or biased. Current approaches evaluate content statistically to identify specific items that are inappropriate because they unfairly favor one group over another. Methods are commonly applied as well to determine whether different constructs are measured across nominal groups by the same test; a test may measure verbal ability in Caucasians, for example, but may be measuring something quite
different (such as exposure to American culture) in a Hispanic American population (Reynolds, 2000; Reynolds, Lowe, & Saenz, 1999). These are only a few of the recent, more sophisticated, methods of considering bias assessment; see Reynolds (2000).

When a test measures something different from what it was intended to measure for specific groups, then it may be considered a biased instrument against those groups. A measure is fair to the extent that the score only includes (a) variables associated with the construct being measured, and (b) random variance from error. In other words, a measure is fair to the extent that it minimizes systematic error in true score estimation as a function of group membership. If a test systematically assigns certain groups lower scores than their “true” score, then the test should be considered biased (Mackintosh, 1998).

This concept is related to the idea of differential validity, in which a test may demonstrate more or less validity for different subgroups (such as ethnicity). There have been a series of studies, all sponsored by the College Board, that have shown differential validity on the SATs. Ramist, Lewis, and McCamley-Jenkins (1994) found that the SATs were correlated higher with first-year college GPA (FYCGPA) for Caucasians ($r = .52$) than for Hispanic Americans ($r = .30$). Bridgeman, McCamley-Jenkins, and Ervin (2000) found that SATs overpredicted African American FYCGPA, particularly for African American males. Finally and more recently, Mattern, Patterson, Shaw, Kobrin, and Barbuti (2008) studied differential validity in the revised SATs. They found that the SATs most accurately predicted Caucasian FYCGPA. The SATs underpredicted Asian Americans FYCGPAs and overpredicted African American FYCGPAs. When high school GPA was added to SAT scores, differential validity dropped significantly.

Indeed, as discussed, one possible way to work toward lowering differential validity and increasing nonbiased assessment would be to administer additional measures (beyond high school GPA) that may tap into other aspects of ability that are not assessed by typical components of the standard examinations that are commonly used for college and graduate school admissions.

### The Importance of Creativity

Creativity itself is an essential part of the human experience. Creativity has been described as the most important economic resource of the 21st Century (Florida, 2002). People who are creative are more likely to have better physical health (Lepore & Smyth, 2002; Pennebaker, 1997; Pennebaker, Kiecolt-Glaser, & Glaser, 1988), and a higher state of general well-being (Carson, Bittner, Cameron, Brown, & Meyer, 1994; Plucker, Beghetto, & Dow, 2004; Richards, 2007; Runco & Richards, 1998). Creativity has been found to increase social harmony (King & Pope, 1999; Russ, 1998), improve mood (Amabile, Barsade, Mueller, & Staw, 2005; Schere, 1998), and reduce personal stress (Nicol & Long, 1996).

Beyond its own importance, however, creativity can also aid overall nonbiased assessment. If measures of creativity, for example, are used to create a fuller picture of an applicant, and if creativity is an unmeasured component of ability, then these measures might also help minimize errors in decision-making regarding such students. This possibility for reduction in error can be conceptualized in two very different and complementary ways.

Assume one is trying to understand a person’s overall cognitive abilities, but only a narrow range of relevant abilities (e.g., those measured by traditional IQ tests) are being assessed. Even if these abilities are being assessed fairly, the ability of the examiner to understand the person’s cognitive functioning may be limited. This limitation may be different for different groups. If, for example, some abilities are more relevant in some cultural settings than others, then the unmeasured abilities may have a differential impact on understanding people’s abilities based on their culture. Such a discrepancy would create bias in the assessment process. To use a cross-discipline example, imagine a test of visual artistic ability that only looked at watercolors. The artistry test would miss many other core components (such as sketching or drawing) and would overreward people who were especially good at watercoloring. In doing so, the test would have an implicit bias against people who were colorblind.

In addition, consider a college admissions program (or a clinical test examiner) trying to assess a limited range of a person’s cognitive abilities (e.g., IQ). If the available tests are believed to be biased, it may be possible to correct some of the systematic error present if other cognitive abilities are tested, as long as these abilities are believed to influence scores adversely on other tests in the examination. If, for example, scores on a valid and reliable test of creativity could be shown to correlate with some systematic ethnic bias of IQ tests, then perhaps the creativity test scores could be used to attenuate any systematic error on the IQ tests. Consider, for example, a young man, Jacob, taking a test comparable to the Similarities subtest on the Wechsler Adult Intelligence Scale-IV (WAIS-IV; D. Wechsler, 2008). Jacob is asked, “What do a dodo bird and a dinosaur have in common?” A 2-point response would likely be, “They are both extinct.” A 1-point response would be, “They are both animals.” Yet if Jacob answered in a completely different (yet appropriate) way, such as “They were both characters in Disney movies” (Alice in Wonderland and Toy Story), he would earn a score of 0.

Creative and accomplished people have commented on this exact phenomenon. Daniel Tammet, a New York Times bestselling author, told an interviewer, “You can imagine all kinds of wonderfully poetic, interesting, creative responses to that kind of question but if it’s not simply and bluntly that they’re both living things, something that is so trivially true that most people wouldn’t even think of it...they would be marked down if they didn’t say it as the answer. The other answer, which is more interesting and creative than the actual response expected, gets zero marks” (quoted in S. Kaufman, 2009).

The lack of creativity on standardized IQ and ability tests is noticed by those not in an academic community. Layperson conceptions of intelligence typically include problem solving and verbal fluency as key components, both of which are associated with creativity (Sternberg, 1985; Sternberg, Conway, Ketron, & Bernstein, 1981). Twelve of the top 50 liberal arts colleges do not

---

1 Although the SATs and GREs are not intended to measure ability, studies have found that both the ACT (Koenig, Frey, & Detterman, 2008) and the SATs (Frey & Detterman, 2004) are strongly related to intellectual ability. These terms will not be used interchangeably, but research on ability tests will be used to discuss standardized college admissions tests for this reason.
require the SAT (Bruno, 2006), with at least one college, Bates, citing as a reason that students who do not submit SAT scores are more likely to pursue original and creative activities (Bates College Office of Communications & Media Relations, 2004). The lack of intellectual breadth in admissions tests causes reduced face validity. Group differences on admissions tests can also reduce face validity.

Creativity can help reduce bias in the standardized assessments used for school admissions in both the “psychometric” and the “layperson” ways. Although the psychometric definition of biased assessment is the most relevant for an academic audience, creativity tests can also increase people’s perceptions of bias in admission assessments. Layperson bias issues will be discussed first. Although there are extensive reviews of cross-cultural research in creativity (e.g., Lubart, in press), this paper will focus on ethnic differences with an American population.

Ethnic Differences in Creativity

Divergent Thinking Tests

Before exploring ethnic differences (or the lack thereof) in creativity, it is essential to distinguish among the different ways that creativity is measured. The most common measures are divergent thinking tests, such as the Torrance Tests of Creative Thinking (Torrance, 1966, 2008), which are based on the early works of pioneers in the field of creativity, such as Guilford (1967). These tests traditionally ask questions or present prompts that can answered in an open-ended fashion. In verbal subtests, a typical question might be “What would happen if we didn’t need sleep?” In figural subtests, one might asked to finish an incomplete drawing and assign a title. Responses are then scored on such dimensions as fluency (how many responses are produced) and originality (how unique are the responses).

Most studies of ethnic differences in creativity have used divergent thinking tests. Such usage is unsurprising; these tests continue to be the most popular creativity assessments, especially in deciding who gets into gifted programs (Hunsaker & Callahan, 1995). Their extreme popularity belies their many flaws, ranging from major statistical issues (Clapham, 2004; Heausler & Thompson, 1988) to their focus on a domain-general view of creativity, a perspective no longer accepted by most creativity researchers (e.g., J. Kaufman & Baer, 2004). Whether scores on the TTCT can be extrapolated to actual creative performance in college (or beyond) is possible (Torrance & Saftier, 1989) but not a given assumption.

That said, researchers have found few differences between African Americans and Caucasians on the TTCT and similar divergent thinking measures (such as the Unusual Uses Test; Guilford, 1967). Although most studies have focused on children and are less relevant to this paper (e.g., Glover, 1976b; Iscoe & Pierce-Jones, 1964; Kalsounis, 1974; Knox & Glover, 1978), other studies have found no differences between African American and Caucasian college students (Glover, 1976a).

A few investigations found some (small) differences favoring African Americans; unfortunately, these studies have only focused on children. Torrance (1971, 1973) found that African American children scored higher on the TTCT than Caucasian children on the Figural tests in fluency, flexibility, and originality; Caucasians scored higher on Figural elaboration and all Verbal subtests. The initial sample compared African American children in Georgia with higher-SES children in Minnesota; when a subsequent study used Caucasians, also from Georgia, all differences were significantly reduced. Kalsounis (1974) also found that African American children received higher fluency and originality scores on the TTCT. Troiano and Bracken (1983) administered measures of creative thinking to three kindergarten classes (Dutch Americans, African Americans, and Native Americans). They found that African Americans and Native Americans scored approximately one standard deviation higher on creative thinking, particularly in fluency, than the Dutch Americans. All of these studies have focused on children, so any extrapolations to college students can only be done very cautiously.

Price-Williams and Ramirez (1977) found an interesting ethnicity by gender interaction, albeit also in a sample of children. African American males (Cohen’s $d = 0.72$) and Hispanic American males (Cohen’s $d = 0.65$) outperformed Caucasian males on the fluency subscale of the Unusual Uses Test (results for flexibility were slightly in favor of African American and Hispanic American males). The results were reversed for females. Caucasians outperformed African Americans (Cohen’s $d = 0.91$) and Hispanic Americans (Cohen’s $d = 0.19$) on fluency. Results were comparable for flexibility (Cohen’s $d$ of 0.89 and 0.20 for African Americans and Hispanic Americans, respectively). There have been few other studies that have found this type of interaction; the vast majority of studies that examine gender differences either find no significant differences or mixed results (Baer & Kaufman, 2008).

Studies of creativity using divergent thinking instruments in Hispanic Americans and Caucasians tend to find different results depending on whether the creativity measure is verbal or nonverbal. For example, Argulewicz and Kush (1984) found that Caucasians scored higher than Hispanic Americans on three of four TTCT Verbal forms, but found no significant differences on the Figural forms. Further, low-income Hispanic American elementary students scored below the norms on the TTCT (Mitchell, 1988). It is worth pointing out, however, that the TTCT has been translated into Spanish (among many other languages), and has been shown to have validity in many Hispanic cultures (e.g., S. Wechsler, 2006).

Some studies have indicated the bilingual students may have an advantage in creative abilities (see Garcia, 2003, for a review). Ramirez and Castaneda (1974) proposed that people who are bilingual have more “cognitive flexibility,” which allows them to approach problems from multiple perspectives. Price-Williams and Ramirez (1977) found that bilingual Hispanic American males outperformed their Caucasian counterparts on the Unusual Uses Test, with Hispanic American females scoring slightly lower than Caucasian females. Carringer (1974) studied Hispanic American children and found that “balanced” bilinguals (those who speak each language equally well) outperformed monolinguals on the TTCT; Kessler and Quinn (1987) found similar results in a group of Hispanic American children.

There are many studies on TTCT differences in Asians and Europeans or Asians and Americans. Zha, Walczyk, Griffith-Ross, Tobacyk, and Walczyk (2006), for example, found that although Chinese graduate students outperformed their American counterparts on the GRE, American graduate students scored higher on four out of five measures of divergent thinking. Yet very few
studies have compared Asian Americans to Americans of different ethnicities on divergent thinking tasks. Yoon (2006) gave the TTCT to European American and Asian American middle school students (the latter being a mix of Chinese American, Korean American, Japanese American, and Southeastern Asian Americans). There were no significant differences either between the European American and Asian Americans or between the different subgroups of Asian Americans.

Creative Performance and Ratings

Beyond divergent thinking tests, one of the most commonly used ways to evaluate creative responses is the Consensual Assessment Technique (CAT; see Amabile, 1982, 1996; Baer, Kaufman, & Gentile, 2004). With the CAT, participants are asked to create something, the creativity of which experts are then asked to evaluate. In the CAT, rather measuring some skill that is theoretically linked to creativity, actual creativity is assessed. It has been used in many studies with a wide range of tasks (e.g., writing poems and stories, telling stories to go with pictures, creating collages and other artworks, and creating mathematical word problems and puzzles) with both children and adults as subjects. In study after study, these expert ratings, done completely independently of one another and without rubrics of any kind, have yielded quite satisfactory interrater reliabilities that typically exceed .70 and often exceed .90 (e.g., Amabile, 1982, 1996; Baer, 1993, 1997, 1998; J. Kaufman, Baer, & Gentile, 2004; Runco, 1989). Indeed, this technique is widely considered one of the best possible creativity assessments (J. Kaufman, Plucker, & Baer, 2008).

Generally, the studies that have used the CAT have found little or no ethnic differences. J. Kaufman et al. (2004) studied poems, stories, and personal narratives written by African American, Hispanic American, Asian American, and Caucasian 8th grade students. There were no differences across ethnicity in creativity scores assigned by expert judges (creative writers, creative writing teachers, and psychologists studying creativity, all of whom showed strong agreement across groups; see Baer, Kaufman, & Rigg, 2009). In a similar study, J. Kaufman, Niu, Sexton, and Cole (2010) looked at novice ratings of poems and stories. They found no ethnic differences among Caucasian, African American, Hispanic American, and Asian American poems or stories. They did, however, find a curious result for rated stories. Both Caucasian and African American novice raters preferred stories written by Caucasians (as opposed to any other ethnicity), even though there were no differences overall (i.e., including ratings by Hispanic Americans and Asian Americans).

Results are less clear for Asian Americans; as with TTCT studies, most of the research is cross-cultural and does not focus on Asian Americans. Artwork produced by American college students, for example, was rated as more creative than art produced by Chinese students by both American and Chinese raters (Niu & Sternberg, 2001). Yet, a similar study that compared American and Chinese drawings of geometric shapes found that the two groups were rated similarly for creativity by both American and Chinese raters (Chen et al., 2002). There were no differences in rated artwork between Chinese and British schoolchildren, except for the higher ratings earned by Chinese children who attended a weekend art school (Cox, Perara, & Fan, 1998). Another study found that Japanese children produced higher rated drawings than British children (Cox, Koyasu, Hirunuma, & Perara, 2001).

Looking specifically at Asian Americans, Niu and Sternberg (2003) compared the creative performance between Asian Americans and Non-Asian Americans, and found no difference between the two groups. Rostan, Pariser, and Gruber (2002) studied Chinese American and Caucasian students’ artwork, with two groups in each culture: students with additional art training and classes and students with no such classes. Each group’s artwork (one drawing from life and one drawing from imagination) was judged by both Chinese and American judges. There were no significant differences between cultures from either set of judges. The only differences found were that art students (regardless of ethnicity) received higher ratings for their artwork than did nonart students.

Paletz and Peng (2009) conducted an investigation of how dialectical styles interacted with ethnicity task type, with creativity as the outcome variable. They found that Asian American and Caucasian college students did not differ significantly on a problem construction task that was assessed by trained coders (as designed by Reiter-Palmon, Mumford, O’Connor, Boes, & Runco, 1997).

Cheng, Sanchez-Burks, and Lee (2008) examined the creativity of Asian Americans as it related to identity integration, the degree to which a person feels that their disparate social identities (such as being Asian and being American) are complementary. They used a cooking simulation in one part of their study in which participants were presented with Asian-only, American-only, or Asian and American possible ingredients. When presented with stimuli from multiple cultures, Asian Americans with high identity integration were more creative (Cohen’s d = 0.56); this finding disappeared in the Asian-only or American-only settings. These findings can be seen as comparable to the findings discussed earlier about balanced bilinguals and acculturated Hispanic Americans demonstrating more creativity.

A comparable technique to asking experts to rate creative performance is to have teachers rate students on their creative abilities. Pfeiffer and Jarosewich (2007) looked at teacher ratings of gifted children on the Gifted Rating Scales-School Form (GRS; Pfeiffer & Jarosewich, 2003); there were no differences by ethnicity (African American, Asian American, Caucasian, and Hispanic American) on either the Creativity or Artistic Talent scales. These findings echo an earlier study by Harty, Adkins, and Sherwood (1984), who studied ethnic differences (African American and Caucasian) across several measures of giftedness. Teacher ratings of creativity were one of the few measures to show no differences by ethnicity. In contrast, however, another study found that teachers rated Caucasian students as being more creative than Hispanic American students, with highly acculturated Hispanic Americans receiving higher marks than less acculturated Hispanic Americans (Masten, Plata, Wenglar, & Thedford, 1999).

Another way to evaluate creative performance is to look at how people perform creative tasks in teams. McLeod, Lobel, and Cox (1996) compared homogenous groups (all Caucasian) to ethnically diverse groups (including African Americans, Hispanic Americans, and Asian Americans) on a brainstorming task. The heterogeneous groups produced ideas rated as more effective and feasible. Milliken and Martins (1996) also found that ethnic diversity was linked with higher idea quality in group tasks. Cady and Valentine (1999) looked at both ethnic and gender diversity and how they related to quality of ideas generated during a brainstorm-
ing task. They found that increased ethnic diversity was associated with a higher quality of ideas ($B = 13.57, \text{eta}^2 = .12, p < .05$), yet increased gender diversity was associated with a lower quality of ideas ($B = -18.32, \text{eta}^2 = .15, p < .01$).

Paletz, Peng, Erez, and Maslach (2004) studied three-person heterogeneous teams that were either primarily Caucasian or primarily minority (most minority participants were Asian American). They asked the teams to create an ending to an ambiguous story; the endings were then rated by appropriate experts. The rated creativity did not significantly differ based on the dominant membership of the group (Caucasian, Asian American, or other ethnicity). It is interesting to note that groups predominantly comprised of minorities reported enjoying the task more. Chatman, Polzer, Barsade, and Neale (1998) found that demographic diversity in creative groupwork helped the most when organizational culture was more collectivistic than individualistic (which, again, is consistent with the research on bilingualism and acculturation).2

**Self Assessments, Styles, and Values**

A third way of broadly measuring creativity is simply by asking people for their evaluations and opinions about their own creativity. Although such creative self assessments often do not correlate to actual performance-based measures of creativity (see, e.g., J. Kaufman, Plucker, et al., 2008), they do provide an indicator of how people view and value their own creativity. J. Kaufman (2006) asked 3,553 individuals (mostly high school and college students) to rate themselves in 56 different domains across five factors (social-communications, visual-artistic, verbal-artistic, science-analytic, and sports). He found that African Americans and Native Americans rated themselves as being more creative than Caucasians, Hispanic Americans, and Asian Americans in science-analytic. Caucasian, African American, Hispanic American, and people of mixed ethnicity ranked themselves higher on the social-communications factor than did Asian Americans. African Americans rated themselves higher than all other ethnicities on both visual and verbal art and higher than Caucasian and Asian Americans on sports.

One notable trend is that African Americans were less likely to fall prone to gender stereotypes than were other ethnic groups. The visual-artistic factor, for example, could be considered to be stereotypically female and was generally rated higher by women in the study than men, yet African American men rated themselves as significantly more creative in these areas than most other ethnicities. The science-analytic and sports factors are stereotypically male and were generally rated higher by men than women. Yet, African American women rated themselves higher than did most other ethnicities.

Other studies have focused on general perceptions of creativity. Some researchers found that Caucasian parents had more favorable perceptions of creativity than Hispanic American parents (Strom & Johnson, 1989; Strom, Johnson, Strom, & Strom, 1992). However, they also found that Hispanic American parents were more likely to engage in play activities with their children and valued play more (Strom & Johnson, 1989). Make-believe play can be a valuable component of a child’s developing imagination (Singer & Singer, 1990).

Other investigations have examined values and self-reported accomplishments. Robinson and Betz (2008) studied work-related values using Super’s Work Values Inventory—Revised (SWVI-R; Zytowski, 2004). The value of “Creativity” showed no differences among Caucasians, African Americans, and Asian Americans. Stricker, Rock, and Bennett (2001) used questionnaires measuring self-reported creative accomplishments and found no difference in college students across all ethnicities.

Another way of examining self-perceptions of creativity is to look at personality tests. The Openness to Experience factor, for example, is highly correlated with creativity across multiple measures of creativity, including self-reports of creative acts (Griffin & McDermott, 1998), creativity ratings on both stories (Wolfradt & Pretz, 2001) and autobiographical essays (Dollinger & Clancy, 1993), and divergent thinking tests (Furnham & Bachtar, 2008; McCrae, 1987). This relationship has been so consistently demonstrated that creative personality tests based heavily on Openness to Experience (e.g., Goldberg et al., 2006) have been used as a proxy measure of creativity (Baer & Oldham, 2006; Powers & Kaufman, 2004).

Unlike the creativity studies I’ve mentioned, there are large-scale personality studies with thousands and thousands of participants (typically adults). There generally tend to be no meaningful differences on any personality factors across cultures (e.g., Goldberg, Sweeney, Merenda, & Hughes, 1998; Kylønen, Walters, & Kaufman, 2005; McCrae & Costa, 1997). However, Heuchert, Parker, Stumpf, and Myburgh (2000) found that White South Africans scored higher on Openness to Experience than Black South Africans (much of this difference was in the Openness to Feelings subcomponent). Alik and McCrae (2004) found that people from European and White cultures tended to be more Open to Experience than people from Asian and African cultures. D. Schmitt, Allik, McCrae, and Benet-Martínez (2007), in a massive study of 17,837 people from 56 nations, found that people from South American and European countries were the most Open to Experience (Chile was the highest), with people from South Asian countries generally being less Open to Experience. African countries were in the middle. It is worth highlighting, however, that Saucier and Goldberg (2001) studied personality labels in 13 languages (including English). They found that Openness to Experience was the only one of the Big Five personality factors to not be comprised of similar words regardless of language. The factor can focus more on intellect, creativity, or unconventionality, for example, depending on the culture. Openness to Experience, therefore, can be considered a concept more readily accepted in Anglo cultures (Benet-Martínez & Oishi, 2008). A further possible issue is that a recent study (DeYoung, Quilty, & Peterson, 2007) was able to tease apart two facets of Openness to Experience: Intellect and Openness. It would quite interesting to see results on all six factors, not merely the standard five.

There are fewer personality studies on ethnic differences with American populations. Asian Americans have been generally found to be less Open to Experience than Caucasians (Benet-Martínez & Karakıpoglu-Aygın, 2003; Leininger, 2002), yet the same studies found similar differences between first-generation

---

2 Curiously, Goncalo and Sawa (2006) examined individualistic and collectivistic-oriented people; when given instructions to be creative, individualistic people generated both a higher number of ideas and more creative ideas than collectivistic people.
Asian Americans and later-generation Asian Americans. Eap et al. (2008) found similar findings (Cohen’s $d = -0.40$) with the added information that Openness was significantly correlated with acculturation ($r = .25, p < .01$).

There is extensive theoretical work on learning and thinking styles, especially as related to creativity. For example, some researchers have proposed that differences on some IQ or achievement subtests, such as those involved in remembering the details of a story, may show larger African American-Caucasian differences, in part because African Americans approach the task differently (Heath, 1983; see Manley et al., 1998). This theory argues that Caucasians approach the task as the test-makers intended—by trying to memorize as many appropriate details as possible and stick to the presented story; in contrast, African Americans may put more emphasis on telling the story creatively. This difference in style would be consistent with Shade’s (1986; Shade & Edwards, 1987) theory of an African American cognitive style, stating that African Americans would be more likely to be spontaneous, flexible, and open-minded. If this theory is true, then African American test performance may suffer because their answers are different from the intended “correct” response. As discussed earlier, such a difference in item interpretation could lead to test bias.

Another possibility to be considered is the phenomenon of stereotype threat. Many studies have suggested that individuals feel stress when placed in a situation where they run the risk of confirming a negative stereotype about their group (e.g., ethnicity). This stress often causes poor performance (Steele, 1997; Steele & Aronson, 1995). Nguyen and Ryan (2008) conducted a meta-analysis of stereotype threat studies found an overall effect size of 0.26. Stereotypes about intelligence are widely known, even among people who are targets of the stereotypes and who do not endorse them (Devine, 1989). As a result, for example, an African American test-taker may worry about confirming negative stereotypes, which causes added stress—and, by extension, a lower performance on the test. Schmader and Johns (2003) argue that stereotype threat causes reduced working memory.

What may then occur is that members of ethnicities who have traditionally scored lower on IQ tests may experience “disidentification” in this domain—in other words, they gradually remove this domain (in this case, analytic and other “IQ test” type abilities) from their conception of self (Crocker & Major, 1989; Steele, 1997). Instead of identifying themselves with these types of abilities, some people may instead identify themselves with other important cognitive abilities that are not associated with IQ tests. One of these abilities, judging from studies like J. Kaufman (2006), might be creativity.

Some studies have, indeed, demonstrated how minorities can use creativity to reduce stress and improve self concept. Metzl (2009) studied survivors of Hurricane Katrina. He found that for all survivors, originality (as measured by the TTCT) indirectly was linked to positive well-being. In African Americans, however, both flexibility and originality were directly linked to both lower stress levels and higher life satisfaction. In addition, survivors of all ethnicities viewed creativity as important factors in their resilience. Similarly, Shinmar (2008) studied how Mexican immigrants best coped with negative social identity. She found that one of the most-used and best-used strategies was social creativity. In this concept of social creativity (e.g., Turner & Brown, 1978), a person changes his or perceptions instead of taking individual action. Shinmar found, for example, that some of her interviewees focused on positive elements of what their jobs meant to them (such as being a hard worker) instead of the negative elements often imposed by others (such as working in a low-status job).

**Creativity as an Unmeasured Component of Intelligence**

Adding creativity as a supplement to the admissions process might combat layperson questions of fairness if these measures showed the same reduced or nonexistent differences by ethnicity that have been found in past research. Creativity would only reduce psychometric notions of bias, however, if the construct of creativity is considered to be part of the abilities that current tests seek to measure. If creativity is orthogonal to current conceptions of intelligence then adding such measures would only obscure any mission to assess intellectual potential or capacity.

Creativity is an important, if frequently nonessential, part of most major theories of intelligence. Guilford’s (1967) Structure of Intellect model, for example, includes divergent production as one of four “operations.” Divergent production was defined by Guilford as the ability to produce information from memory to meet specific objectives. The construct was conceived “through investigations of certain hypotheses regarding the component abilities most relevant to creative performance” (p. 169). This work, followed up by other researchers (most notably Torrance, 2008, as discussed), has often been used as a measure of creativity.

Many subsequent theories of intelligence incorporated creativity-related abilities into their structures. Undoubtedly, the theory of intelligence that is most often applied to IQ tests is the CHC (Cattell-Horn-Carroll) theory, a combination of two earlier theories. The Cattell-Horn theory (e.g., Horn & Cattell, 1966) initially proposed two types of intelligence, crystallized ($G_c$) and fluid ($G_f$). $G_c$ signifies what a person knows and has learned, and $G_f$ represents how a person handles a new and different situation (i.e., novel problem solving). Horn (1985) expanded the theory to include more dimensions (known as Broad Abilities), such as Visual Processing ($G_v$), Short-Term Memory ($G_{sm}$), and Processing Speed ($G_s$). Carroll’s (1993) theory proposed a hierarchy of intellectual abilities. At the top of the hierarchy is general ability; in the middle of the hierarchy are various broad abilities, similar to Horn’s (including learning and memory processes and the effort-less production of many ideas). At the bottom of the hierarchy are many narrow, specific abilities, such as spelling ability and reasoning speed (creativity was placed as a narrow ability).

The combined CHC theory incorporates the concept of a general intelligence in that all of the different aspects of intelligence are considered to be related to a common “$g$” (although this aspect is not often emphasized and was not at all advocated by Horn; see Flanagan & Ortiz, 2002). The theory also includes the concept of many different aspects of intelligence. Although in the early stages of the Cattell-Horn $G_c-G_f$ theory, $G_f$ was hypothesized to be strongly linked to creativity (Cattell & Butcher, 1968), such a relationship is no longer explicitly part of the CHC theory. The current model, based in large part on factor analytic studies by Carroll (1993) and others, includes originality/creativity as a component of long-term storage and retrieval ($G_c$). According to the most recent presentation of CHC theory (McGrew, 2009), “Some
Glr narrow abilities have been prominent in creativity research (e.g., production, ideational fluency, or associative fluency)” (p. 6). In the detailed description of the model, this sentence is the only mention of creativity, originality, or divergent thinking. Gf is discussed in terms of its relationship to problem-solving and coping with novel problems (both considered to be highly related to creativity), yet the emphasis is on Gc.

The PASS (Planning, Attention, Simultaneous, and Successive) theory is a cognitive processing theory based on the works of Luria (see Das, Naglieri, & Kirby, 1994, for an overview). Like the CHC model, the PASS model is used as a conceptual framework for intelligence tests. Luria’s (1966, 1970, 1973) original neuropsychological model featured three Blocks or functional units. The first unit is responsible for focused and sustained attention. The second functional unit receives and stores information and features both simultaneous and successive (or sequential) processing. Simultaneous processing involves integrating chunks of information together, largely in parallel; chunks are synthesized together simultaneously, much as one might appreciate a painting all at once. Successive processing is interpreting chunks of information separately, in sequential fashion, much as when one listens to a long string of directions told to you by a tollbooth operator and then you have to remember those directions exactly to arrive at your destination. The third functional unit is responsible for planning, decision-making, and self-monitoring behavior. It is this last process, planning, that has been hypothesized to be most related to creativity (Naglieri & Kaufman, 2001). For example, one study found that people who spent time planning and replanning a project were more productive and more creative than those who spent less time planning (Redmund, Mumford, & Teach, 1993).

One theory of intelligence that specifically includes creativity as a key component is Sternberg’s (1996, 1999; Sternberg et al., 2008) theory of successful intelligence. This theory comprises three “subtheories”: a componential subtheory, which relates intelligence to the internal world of the individual; an experiential subtheory, which relates intelligence to both the external and the internal worlds of the individual; and a contextual subtheory, which relates intelligence to the external world of the individual. The componential subtheory specifies the mental mechanisms responsible for planning, carrying out, and evaluating intelligent behavior. The experiential subtheory expands on this definition by focusing on those important behaviors that involve either adjustment to relative novelty, automatization of information processing, or both. The contextual subtheory defines intelligent behavior as involving purposeful adaptation to, selection of, and shaping of real-world environments relevant to one’s life (Sternberg et al., 2008). Creativity encompasses many of the abilities needed for the experiential subtheory (i.e., reacting to new and different stimuli).

Gardner’s (1999) well-known theory of multiple intelligences does not specifically address creativity. However, his eight intelligences (interpersonal, intrapersonal, spatial, naturalistic, linguistic, logical-mathematical, bodily kinesthetic, and musical) certainly pertain to creativity. Gardner (1993) used case studies of eminent creative individuals to argue that creative people can shine as a function of embodying different intelligences. For example, he selected Freud as an example of intrapersonal intelligence; Einstein to represent logical-mathematical intelligence; Picasso, spatial intelligence; Stravinsky, musical intelligence; T. S. Eliot, linguistic intelligence; Martha Graham, bodily kinesthetic intelligence; and Gandhi, interpersonal intelligence (naturalistic intelligence had not been added at this time).

Although creativity is an aspect of the CHC and Luria-based theories, it is not measured (nor is it claimed to be measured) on any of the IQ tests developed from one or both of these theories, such as the Wechsler Intelligence Scale for Children IV (D. Wechsler, 2003), Kaufman Assessment Battery for Children-II (A. Kaufman & Kaufman, 2004), Woodcock-Johnson-III Tests of Cognitive Ability (Woodcock, McGrew, & Mather, 2001), or the Stanford-Binet Intelligence Scales, Fifth Edition (Roid, 2003). Conversely, neither Sternberg’s theory of successful intelligence nor Gardner’s multiple intelligences have been the basis of an IQ test or a traditional college admissions test. Sternberg’s theories have been the basis for an alternative college admissions tool, which will be discussed later.

Even if IQ tests and the SATs/GREs were perfectly equitable, ability assessment could still include some bias. Creativity is generally considered a part of intellectual ability, yet it is not reflected in any major measurement used for college admissions. Further, given the widespread lack of ethnic differences, there is evidence that this absence may be detrimental to minorities.

### The Role Creativity Might Play in Admissions

As has been discussed, creativity might reduce bias in college admission assessments in two distinct ways—by having fewer (or no) differences across ethnicity (the layperson approach to bias) and by supplementing the current ability and achievement measures to better measure the construct of intellectual ability (the psychometric approach). Despite this possibility, little actual research has studied how creativity might impact admissions.

One major exception can be found in the work of Sternberg and his colleagues. They studied how college admissions would be impacted by supplemental measures of successful intelligence (such as practical intelligence and creativity). The assessments used are primarily based on Sternberg’s work on open-ended creativity assessments. Specifically, Sternberg and colleagues (Sternberg, 2008; Sternberg & the Rainbow Project Collaborators, 2006) used three types of open-ended measures. Students were asked to write captions to a cartoon, write short stories based on titles like “The Octopus’s Sneakers,” and tell stories based on a selected series of images. These responses were then evaluated by trained judges for cleverness, humor, originality, and task appropriateness (for the cartoons), and originality, complexity, emotional evocativeness, and descriptiveness for both written and oral stories (Sternberg & Lubart, 1996; Sternberg & the Rainbow Project Collaborators, 2006). This procedure is quite similar in nature to the methodology used for the Consensual Assessment Technique. Related projects created “augmented” versions of the AP Test for Psychology, Statistics, and Physics that assessed the information via analytic, creative, practical, and memory-based subscales (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Stemler, Sternberg, Grigorenko, Jarvin, & Sharpes, 2009). This first phase of the project, conducted at many different universities, was called the “Rainbow Project.”

Some of this work was done using the actual admissions criteria for Tufts University. Sternberg dubbed his work with these new measures (including creativity) at Tufts University the “Kaleidoscope Project.” Similar to earlier work, students write stories (with
titles like "Confessions of a Middle School Bully"), describe an alternate history in which Rosa Parks gave up her seat on the bus, and design and advertise new products. These measures were included (optionally) in the 2006–2007 application for the 15,000 students who applied to arts, sciences, and engineering at Tufts.

The results from Sternberg’s earlier work and his Tufts work at Tufts are encouraging. His measures (not only of creativity, but for all components of successful intelligence) predict college success more accurately than standard admissions tests (with twice as much predictive power than SATs alone; Sternberg, 2008). In addition, and most relevant for the purposes of this discussion, ethnic differences were significantly reduced (Sternberg, 2006; Sternberg & the Rainbow Project Collaborators, 2006). For the Rainbow Project, creativity scores on a multiple choice measure correlated .35 with first year college GPA. Creative performance measures correlated .29 (for oral stories), .12 (for written stories), and .08 (for cartoon caption-writing) with first year college GPA. Only cartoon caption-writing was not significant. Given that there are ethnic differences on the SATs, which predict first year college GPA quite well (Sackett et al., 2008), it is interesting to note that the differences between Caucasians and other ethnicities were much smaller on the creativity measures than on the SATs. For written stories, for example, the Cohen’s d scores, using Caucasians as the reference group, were −0.26 for African Americans, −0.11 for Hispanic Americans, −0.25 for Asian Americans, and 0.01 for Native Americans. For the most relevant SAT score (Verbal), the comparable Cohen’s d scores were −0.73 for African Americans, −1.10 for Hispanic Americans, −0.23 for Asian Americans, and −0.62 for Native Americans. Cartoons and oral stories showed similar patterns; no Cohen’s d score was below −0.51, and Native Americans outperformed Caucasians on oral stories with 0.50 (Sternberg, 2008).

For the AP test investigations, as with the SAT studies, ethnic differences were also reduced. In the actual AP Psychology test, for example, African American studies have Cohen’s d scores of −0.72 compared to Caucasians. The augmented test reduced differences (Cohen’s d = −0.24); the creative scale reduced differences to be nearly nonexistent (d = −0.02). For Hispanic Americans, the actual test show high differences (d = −0.58), yet the augmented test (d = −0.47) and creative scale (d = −0.32) are substantially lower (Stemler et al., 2006). Similar results were found on the AP Statistics and Physics tests. On the statistics tests, African American Cohen’s d scores went from −1.10 (augmented test) to −0.30 (creativity scale); for Hispanic Americans, the d scores were −0.59 for the augmented test and −0.47 for the creativity scale (Stemler et al., 2006). On the physics tests, African American Cohen’s d scores went from −1.05 on the augmented test to −0.55 on the creativity scale; for Hispanic Americans, the d scores were −0.61 for the augmented test and −0.39 for the creativity scale. This study also examined Asian Americans, who were not significantly different from Caucasians on either the augmented test or the creativity scale (Stemler et al., 2009).

At Tufts, the quality of applicants rose (paradoxically, despite de-emphasizing SAT scores, the average SAT scores of applicants increased), and minority admissions went up (Sternberg, 2008). Specifically, in the first year of the program, admissions of African Americans rose 30% and Hispanic Americans rose 15%. At the end of the first year, students with high scores on the Kaleidoscope measures performed comparably to those admitted primarily for other reasons (i.e., SAT scores or high school GPA). Extracurricular and leadership activities were tied to Kaleidoscope scores (Sternberg, 2008). Most notably, there were no significant differences by ethnicity found on any of the Kaleidoscope measures (Sternberg & Coffin, 2010).

Although no other studies have evaluated how creativity tests might impact minority enrollment, other studies have further examined how well creativity might predict or relate to school performance. Altman (1999) administered a series of divergent thinking tests and accessed the participants’ GPA scores. A composite divergent thinking score was significantly correlated to overall GPA (r = .20, p < .01), although there was a bimodal distribution, with the relationship strongest in early courses and very advanced courses. It is worth noting, however, that whereas these tests have accounted for some variance in academic success, other cognitive variables have accounted for a higher percentage of the variance (Niaz, Saud de Nunez, & Ruiz de Pineda, 2000). Similarly, Ai (1999) found that teacher assessments of student creativity were related to academic achievement (as measured by student self-report), but more psychometric tests of creativity showed little relation. Although Toth and Baker (1990) found a trend in the opposite direction (students with high divergent thinking scores and nonconvergent learning styles showed lower levels of classroom achievement), the participants were school-age children.

One study focused on the highly competitive private school Choate Rosemary Hall (Grigorenko et al., 2009). Choate students who performed well on two creativity tasks (writing and science) also had higher GPAs. The creative science task significantly predicted first year GPA, although the creative writing task did not. In addition, a rated creativity score (by admissions officers) was not predictive of first year GPA, either. Grigorenko et al. (2009) argue that the admissions ratings’ lack of predictive validity is a sign that performance-based tests are needed.

At the other end of the educational spectrum, Sternberg and Williams (1997) examined how GRE scores compared to other factors (such as advisor-rated creativity) in predicting graduate student success. They found that the GRE-Verbal (r = .18, p < .05), the now defunct GRE-Analytical (r = .17, p < .05), and the GRE-Psychology (r = .37, p < .01) predicted first year graduate student grades (GRE-Quantitative did not), whereas none of the GRE tests predicted second year grades. Of more interest is that they did not predict the advisor’s rating of a student’s research or teaching potential, and only the GRE-Analytical (r = .24, p < .05) predicted the dissertation reader ratings. In contrast, advisor-rated creativity predicted overall Year One and Two grades (r = .16, p < .05) and advisor-rated research potential (r = .16, p < .05, p < .05, p < .05) and dissertation reader ratings (r = .34, p < .01). It is important to note, however, that the study suffered from a restriction of range problem. Perhaps the extra role that creativity may play in dissertation quality (as opposed to the traditionally used criterion of first year grades) can be explained by Chamorro-Premuzic (2006), who studied how creative thinking and conscientiousness differed in predicting student success. He found that creativity was associated with how well students performed on their dissertations, whereas conscientiousness was more associated with performance on exams. Creative students tended to prefer oral exams, group projects, and
working on their dissertation; more conscientious students preferred multiple choice and essay exams.

N. Schmitt et al. (2009) conducted a large-scale investigation of how cognitive and noncognitive factors predict college success, with an emphasis on the potential impact of noncognitive information on minority applicants. They collected biodata on incoming freshman students that measured several essential noncognitive constructs, such as leadership, interpersonal skills, and perseverance. Unfortunately, creativity was not explicitly measured, although other measured constructs included learning (continuous learning, intellectual interest, and curiosity), as well as artistic and cultural appreciation. Both of these constructs can be assumed to correlate positively with creativity. What N. Schmitt et al. (2009) found was that most of the biodata constructs correlated significantly with cumulative GPA in the students’ senior year, including the learning ($r = .14, p < .05$) and artistic ($r = .19, p < .05$) variables. Most noteworthy was N. Schmitt et al.’s (2009) finding that if the alternate measures (including these creativity-related constructs) had been used to supplement the traditional GPA-standard test admissions policy, then a greater proportion of Hispanic and African American students would have been admitted.

**Standardized Tests and a Small Movement Toward Creativity**

Although many more studies are needed to affirm creativity’s possible use as an admissions tool to reduce bias and promote diversity, it seems to have striking potential. However, it is important to note that the people who make standardized tests are not stupid. If it were easy to add a construct like creativity to a standardized test, it would be done. Researchers at the Educational Testing Service (ETS), which create the GREs and are involved in the creation of the SATs, have been considering measures of creativity for over 50 years (Stewart, 1953). Frederiksen (1959) and colleagues (Frederiksen & Evans, 1974; Frederiksen, Evans, & Ward, 1975) developed and tested a measure of scientific creativity to see how many hypotheses someone could generate in response to a graph or chart. Although ETS still uses this task in research studies (now adapted for computer use), the measure has yet to make it onto the actual test. One reason is likely its extremely low correlation with the rest of the test (Bennett & Rock, 1995).

Creativity simply does not yet have an affordable, valid, and easy-to-administer large-scale measure that is not susceptible to coaching and faking. There is a clearly desired outcome (which school would specifically look for an uncreative student?). Indeed, the ease with which most creativity tests can be faked (e.g., Ironson & Davis, 1979) is a leading reason why many standardized tests avoid the construct altogether (Kyllonen, 2007; Kyllonen, Walters, & Kaufman, 2006). Those tests that are much harder (if not impossible) to fake, such as the Remote Associates Test (Mednick, 1968), are often more heavily dependent on intellectual abilities (such as vocabulary), thereby confounding the results.

ETS would likely argue that their standardized tests are linked to creativity. Indeed, Powers and Kaufman (2004) found that people who were more creative (as measured by a personality-style test) also scored higher on the GRE. Wang (2007) found that CBEST scores (an achievement test for teachers) were strongly related to creativity. Kuncel et al. (2001) used a meta-analysis to argue that the Miller Analogies Test not only predicts academic performance but also creativity. These findings are unsurprising, however, given creativity’s extensively documented (if moderate) link with intellectual abilities (Barron & Harrington, 1981; Kim, 2005). Whether the connection between creativity and standardized test performance represents anything notable beyond a common ground of intellectual abilities has yet to be seen.

ETS has developed a new Personal Profile Indicator (PPI), which is a standardized letter of recommendation for professors and other recommenders to fill out for graduate school applicants (Kyllonen, 2008; Walters, Plante, Kyllonen, Kaufman, & Gallagher, 2004). Six key abilities are highlighted on the PPI, including Knowledge/Creativity. Creativity’s inclusion is unsurprising given past research; when faculty members are asked which factors are important for success in graduate school, creativity invariably is included (Enright & Gitomer, 1989; Powers & Fowles, 2000; Walpole, Burton, Kany, & Jackenthal, 2001).

Grigorenko et al.’s (2009) findings about the predictive power of ratings, however, give one pause. Sternberg and Williams (1997) used ratings from a graduate school advisor, the predictive power of these scores may be because of the advisor’s advanced knowledge of a student. It is questionable how meaningful or valid creativity assessment via the PPI, as rated by high school teachers, will be. It is further questionable how much value schools will grant to a new and untried measure. The PPI, however, remains a step forward (albeit a small one).

**Can Creativity Actually Help?**

Unlike higher education admissions, workplace personnel selection typically values creativity (Burch, Pavelis, & Port, 2008). Often creativity is measured as part of a larger personality assessment, with Openness to Experience being a desired trait (Moy & Lam, 2004; Ployhart, Weekly, Holtz, & Kemp, 2003). Even if other personality attributes are usually valued more (such as conscientiousness; Schmidt & Hunter, 1998), creativity still plays a role. More recently, the selection process has encompassed assessments specifically designed to capture creativity or innovation, such as the Innovation Potential Indicator (IPI; Patterson, 2000) and the Team Selection Inventory (TSI; Anderson & Burch, 2003).

Could creativity assessments be used in higher education the way they are used in the workplace? There are a number of legitimate obstacles. Creativity measurement is notoriously spotty and inconsistent. The most popular creativity tests are based on the concept of divergent thinking and have numerous flaws (Baer, 1994; Clapham, 2004). Other possible creativity measurements, such as the Consensual Assessment Technique, tend to be complicated and expensive (J. Kaufman, Baer, Cole, & Sexton, 2008). Self assessments hold inconsistent relationships to other creativity measures (J. Kaufman, Evans, & Baer, 2010).

The upside, however, is substantial. Although Messick (1980) considered face validity to be less important than other types of validity, he did acknowledge that strong face validity can improve performance on ability tests (Messick, 1995). Atkinson and Geiser (2009), in reviewing the state of college admission tests, argue that face validity needs to be a core component of any test used for admissions. Indeed, Hollis-Sawyer and Sawyer (2008) found that a low stereotype threat condition combined with high test validity produced the highest minority scores (African American and His-
panic American). Most minorities grow to view standardized tests as being biased in favor of Caucasians (McKown & Weinstein, 2003), including arenas of intelligence that do not have these negative associations may be a way of combating issues like stereotype threat.

Indeed, despite creativity’s role in and relation to intelligence, ethnic differences are rare. People generally tend to be creative across all ethnicities, regardless of whether that creativity is measured by older methods (such as divergent thinking tests) or newer methods (such as assessing creative performance). At Tufts, Sternberg’s pioneering work has led to higher minority enrollment while maintaining institutional excellence.

Beyond mere face validity, creativity can possibly impact fairness in admissions in a more sophisticated way. As discussed earlier, a measure is fair to the extent that it minimizes systematic error in true score estimation as a function of group membership. If some ethnic groups score better, comparatively, on Aspect A of intelligence (such as creativity) than Aspect B (such as a more traditional measure, memory), then a test that only measures Aspect B and not Aspect A is biased. It is important to note that the relationship of creativity to intelligence is frequently debated. Some may consider creativity an entirely separate construct (see Sternberg & O’Hara, 1999, for a discussion of possible relationships between creativity and intelligence) and would therefore argue that it is entirely appropriate to omit creativity from intellectual assessment. A steady stream of research over the past 50 years, however, indicates that there is a significant positive relationship between intelligence and creativity (e.g., Kim, 2005).

The possibility of reduced ethnic bias in college admissions assessment should be too tempting to resist for those pursuing this line of research. A comparable analogy can be found in decades-old research on maternal effects on intelligence, specifically the research on identical twins who either shared or didn’t share a placenta (e.g., Rose, Uchida, & Christian, 1981). In brief, twins that shared a placenta were more alike on both verbal and nonverbal measures of intelligence, whereas twins that did not share a placenta were less similar on nonverbal intelligence (for more information, see A. Kaufman, 2009; A. Kaufman & Lichtenberger, 2006). The implications of these findings were huge—a placenta is, in a way, the very first “environment.” If these studies were genuinely explored, it could have produced amazing insights into the relationship between genetics and environment and intellectual ability. Yet although these studies were replicated in France in the mid-1990s, they have been little explored since. Indeed, other scholars in the field have either explained them away or simply ignored them as an aberration. Yet, some scholars (e.g., A. Kaufman, 2004) passionately argued (in vain) that these studies needed be pursued. If these findings might be true, then how dare both researchers and practitioners not consider the possible implications?

I see an analogy with the role of creativity on admissions tests. With a few notable exceptions, no one is pursuing this line of research or trying to implement some of these ideas. Some concepts are too important to be easily dismissed. If—as suggested by Sternberg and his colleagues—minority enrollment could improve by adding creativity to the admissions equation, then it is our duty to investigate further and to make it a mandate to apply the results of this research directly to the college admissions process.

References


Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I*:...


Received March 26, 2010
Revision received March 26, 2010
Accepted May 10, 2010

CREATIVITY AND ETHNIC BIAS


Received March 26, 2010
Revision received March 26, 2010
Accepted May 10, 2010

CREATIVITY AND ETHNIC BIAS


Received March 26, 2010
Revision received March 26, 2010
Accepted May 10, 2010

CREATIVITY AND ETHNIC BIAS


Received March 26, 2010
Revision received March 26, 2010
Accepted May 10, 2010

CREATIVITY AND ETHNIC BIAS


Received March 26, 2010
Revision received March 26, 2010
Accepted May 10, 2010

CREATIVITY AND ETHNIC BIAS