Cognitive Ability Testing and Adverse Impact in Selection: Meta-analytic Evidence of Reductions in Black-White Differences in Ability Test Scores Over Time

Authors
Stephen A. Woods, Claire Hardy, & Yves R. F. Guillaume. Aston Business School, Aston University, Birmingham, UK.
Address for Correspondence: Stephen A. Woods Ph.D. Work and Organisational Psychology Group, Aston Business School, Aston University, Aston Triangle, Birmingham, B4 7ET. S.a.woods@aston.ac.uk
From: Proceedings of the 2011 BPS Division of Occupational Psychology Conference, Stratford, UK.

Summary
Black-White differences in cognitive ability test (CAT) performance were examined meta-analytically in this study. Controlling for sampling effects, meta-analyses revealed a narrowing of differences in CAT performance (on tests of general cognitive ability) between Blacks and Whites over successive decades of the late 20th Century, supporting a cultural/developmental theory of ethnic group differences in test performance, most likely reflecting societal changes in equality of opportunity.

Abstract
Black-White differences in cognitive ability test (CAT) performance were examined meta-analytically in this study. The primary purpose was to test a cultural/developmental theory of differences in CAT performance, which proposes that such differences arise because of educational, social, and economic inequality between ethnic groups. Controlling for sampling effects, meta-analyses revealed a narrowing of differences in CAT performance (on tests of general cognitive ability) between Blacks and Whites over successive decades of the late 20th Century, supporting the cultural/developmental theory. The extent, and rate of this narrowing was moderated by sample type, with differences being smallest, and narrowing fastest, for samples drawn from selective environments (higher education and employed jobs). The most likely explanation of the reduction in Black-White differences is the effect of societal-level changes in attitudes and policies, which have served to promote racial equality and increase opportunity for African Americans.

Main Body
Research into ethnic group differences in cognitive ability, and performance on cognitive tests, is amongst the most controversial and emotive in applied psychology. Meta-analyses have consistently reported systematic differences between White, Black, and Hispanic test-takers in the US, of magnitudes that would almost certainly lead to adverse impact against minority groups in employment testing (Hough et al., 2001). However, recent studies of IQ change, which examine data collected over the past 40 years, suggest that societal changes in opportunities for employment and education may have served to narrow the pervasive gap between Black and White test takers (e.g. Dickens and Flynn, 2006). This could have profound implications for our understanding of the antecedents of ethnic differences in cognitive ability test (CAT) performance, and yet no study has examined this possibility using meta-analyses, representing a major gap in our knowledge of how ethnic differences in cognitive ability may be changing over time.

In this study, we report new meta-analyses that extend previous studies in important ways. Our major contribution is an examination of trends in cognitive ability test performance differences between Blacks and Whites over four decades (1960s, 1970s, 1980s, and 1990s).
In these analyses we controlled for the effects of sampling, comparing samples from specific kinds of sampling environment. We argue for a cultural/developmental theory of CAT performance differences, and demonstrate support for this theory from our meta-analyses.

Explanations for ethnic differences in test performance that posit real differences in the average levels of cognitive ability of ethnic groups are strengthened by inconsistent evidence of test-specific and process explanations. There are two possible theories that explain the antecedents of real ability differences between people from different ethnicities. A distributional theory (e.g. Rushton & Jensen, 2005) suggests that the differences in the CAT scores of people of different ethnicities reflect underlying differences in group characteristics, resulting primarily from ethnicity-specific genotypes. The counterpoint to the distributional approach might be termed a cultural/developmental theory. In this perspective, ethnic differences in cognitive ability test performance are still assumed to reflect real differences in cognitive ability, but these differences are attributed to societal differences in education and opportunity (Dickens & Flynn, 2006; Suzuki & Aronson, 2005). Proponents point out that groups that have traditionally been shown to score lower on CATs are typically overrepresented in poorer and more deprived communities (e.g. Suzuki & Aronson, 2005).

Consistent with a cultural/developmental theory of ethnic differences in CAT performance, we propose that society-level improvements in equality of opportunity and education will have reduced the gap in cognitive ability of Blacks and Whites over successive decades, which should be represented in corresponding narrowing of differences in performance on tests of general ability. The mechanism for raising cognitive ability, and thusly CAT performance of Black test-takers, is most likely to be improvements in education opportunity, affecting ability and test performance directly through learning, and indirectly through reducing disadvantage among communities. In our analyses, we test this theory by examining Black-White differences in CAT performance on a variety of different instruments organized within the last four decades of the 20th Century.

Method

Literature Search

Studies were identified using several literature search strategies. A bibliographic search of electronic databases including PsycINFO, Sciedirect, SwetsWise, Business Source Premier (EBSCO), SocIndex with Full Text (EBSCO), and Ingentaconnect using Boolean searches of keywords including cognitive ability, intelligence, ethnicity, race and adverse impact. A manual search for relevant papers was also conducted in the following journals: Personnel Psychology, Journal of Applied Psychology, Journal of Occupational and Organisational Psychology, and International Journal of Selection and Assessment.

To be included in the present meta-analysis, studies had to meet six criteria. First, individuals in the sample had to be at least 16 years of age. The second criterion was data must be at the individual level and not at the group level. Thirdly, the data had to be in raw form and not transformed in any way (e.g. other-ratings of the test performance). Fourth, no clinical population data were included. Fifth, studies had to contain means and standard deviations, or an appropriate statistic from which to calculate standardized differences (e.g. an F statistic), and report these for specified ethnic groups (i.e. not simply White and ‘Other’). Lastly, cognitive ability tests had to be valid tests of general cognitive ability (g). Tests were
deemed to be valid measures of cognitive ability if: 1) the article specified that the test was an overall test of general mental ability; or 2) the test comprised two or more facets of g (e.g. verbal, math, spatial ability and so forth).

Data Set

Applying the specified inclusion criteria resulted in an initial set of 67 independent samples (n = 803121 individuals) from four decades (1960s, 1970s, 1980s, and 1990s. Following the coding structure of Roth et al. (2001), several specific variables were coded for, including ethnic group, sample type, and level of cognitive ability. We also coded for time (year of data collection).

To code for sample type, we replicated Roth et al’s (2001) approach in which samples were coded as either industrial, military, or educational. We then further subdivided each of the three sample types to reflect the difference between selective and non-selective sampling frames. Military samples were divided into enlisted (selective) versus applicants (non-selective), educational samples into college/university (selective) versus high-school (non-selective), and industrial into incumbent (selective) versus applicants (non-selective). To examine trends in Black-White differences over time, we coded time broken down into decades (i.e. 1960s, 1970s, 1980s, 1990s).

Results

Our analyses examining differences in cognitive ability between Blacks and Whites showed the overall difference for tests of g (defined as those with two or more different test components) was $d = 0.80$ ($K = 91$; $N = 1,110,038$). A significant Q-test ($\chi^2 (91) = 551.40, p < .01$) indicated the presence of moderators. We first looked at sample type as potential moderator, and found the highest differences for the Educational samples ($d = 0.88$), followed by the Military samples ($d = 0.78$), and the Industrial samples ($d = 0.74$).

Significant Q tests for the Educational ($\chi^2 (17) = 174.07, p < .01$), Industrial ($\chi^2 (24) = 124.53, p < .01$), and Military ($\chi^2 (45) = 205.83, p < .01$) subsamples warranted the search for further moderators. Therefore, we next examined the effects of sampling frame by comparing selective versus non-selective sample types in each of the three sample sources, and then by combining the three selective samples (military-enlisted, educational-college/university, and industrial-incumbent), and the three non-selective samples (military-applicant, educational-high school, and industrial applicant). Our results show clear effects of sampling frame in these studies. For the selective samples, the differences were military-enlisted ($d = 0.73$), educational-college/university ($d = 0.75$), industrial-incumbent ($d = 0.54$), and pooled selective sample ($d = 0.71$). For the non-selective samples, differences were markedly higher; military-applicant ($d = 1.31$), educational-high school ($d = 1.27$), industrial-applicant ($d = 0.90$), and pooled non-selective sample ($d = 1.04$).

These results highlighted the importance of including sample characteristics in our comparisons of d over time. We computed $d$ statistics for studies using tests of g, grouped by decade. All of the analyses point to a similar trend, representing an increase in differences in cognitive ability from 1960s to 1970s, with smaller $d$ values in the 1990s. For ease of comparison, we present the data from the combined industrial and educational selective and non-selective samples graphically in Figure 1. The trend in differences across the decades shows decreases in d-scores between 1970s and 1990s. Importantly, our analyses show that
this trend holds when controlling for sample characteristics. Comparing d-scores for similar kinds of samples over time shows a narrowing of differences in CAT performance of Blacks and Whites on tests of g.

Comparing 1970s values with 1990s values for the combined educational and industrial samples, split across selective and non-selective sampling approaches is most illustrative. For selective samples, d-values were 0.92 in the 1970s and 0.62 in the 1990s. For non-selective samples, d-values were 1.16 in the 1970s and 0.81 in the 1990s. Applying Hunter and Schmidt’s (1990) z test revealed that both differences were significant (selective sample: $\Delta d = .30; z = 3.20; p < .01$; non-selective sample: $\Delta d = .35; z = 4.16; p < .01$) These represent a narrowing of mean differences between Blacks and Whites of 0.30 and 0.35 standard deviations respectively, remarkably similar despite the differences in magnitude of d-values between the selective and non-selective samples.

We note the anomalous findings derived from data collected in the 1960s, for which d values were smaller than for the 1970s. This could be explained by considering developments in US society in the 1960s. It is quite plausible that prior to the Civil Rights Act in 1964, African Americans may have been less likely to be granted access to social or employment settings in which testing was prevalent. Those who did may have required an advantage of higher than average levels of cognitive ability. The result would be an effect of range restriction for those groups of participants.

Figure 1. Black-White effect size differences in performance on tests of g, from selective and non-selective educational and industrial samples, across four decades.

<table>
<thead>
<tr>
<th>Year</th>
<th>Selective Samples</th>
<th>Non-selective Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960s</td>
<td>0.92</td>
<td>1.16</td>
</tr>
<tr>
<td>1970s</td>
<td>0.62</td>
<td>0.81</td>
</tr>
<tr>
<td>1980s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

All of the analyses that we ran showed a general reduction in the differences in g between Blacks and Whites from the 1970s to the 1990s. We derived robust evidence of this trend by controlling for sample characteristics. We observed the trend in 1) combined
industrial/educational selective and non-selective samples; 2) industrial applicants and incumbents; 3) educational samples drawn from colleges and universities.

Our findings support a cultural/developmental theory of ethnic differences, in which differences in performance are explained by socio-demographic differences in opportunity, education, and disadvantage. Our results offer a much more optimistic picture than previous meta-analyses and narrative reviews, suggesting that the cognitive ability gap between Blacks and Whites is gradually being eroded over time. Policy makers are warranted to continue to further support and reinforce these developments through equality and affirmative action policies that increase opportunity for minority ethnic groups, and that promote a pro-diversity perspective in organizations. However, given the gap between Blacks and Whites is still larger than half a standard deviation in the most recent data practitioners remain justified in exercising particular care and caution in the use of CATs for selection assessment. The findings also have important implications for how to conceptualise test bias and adverse impact in test design and assessment practice.

References