Adverse Impact & Validity Evidence in South Africa: 12 Years of Data Trends

Prof Hennie Kriek

Managing Director: SHL South Africa

Professor Extraordinarius: UNISA

Ms Kim Dowdeswell

Research Consultant: SHL South Africa

PEOPLE PERFORMANCE

Background to Using Assessments in the South African Context

The Context of Assessment in SA

- History
- Clinical Assessments
- Industry
- Legislation
- Registration of Psychologists
- Bias and Fairness

Racial Group Differences in the South African Population

Group Differences in South Africa: Ability Measures

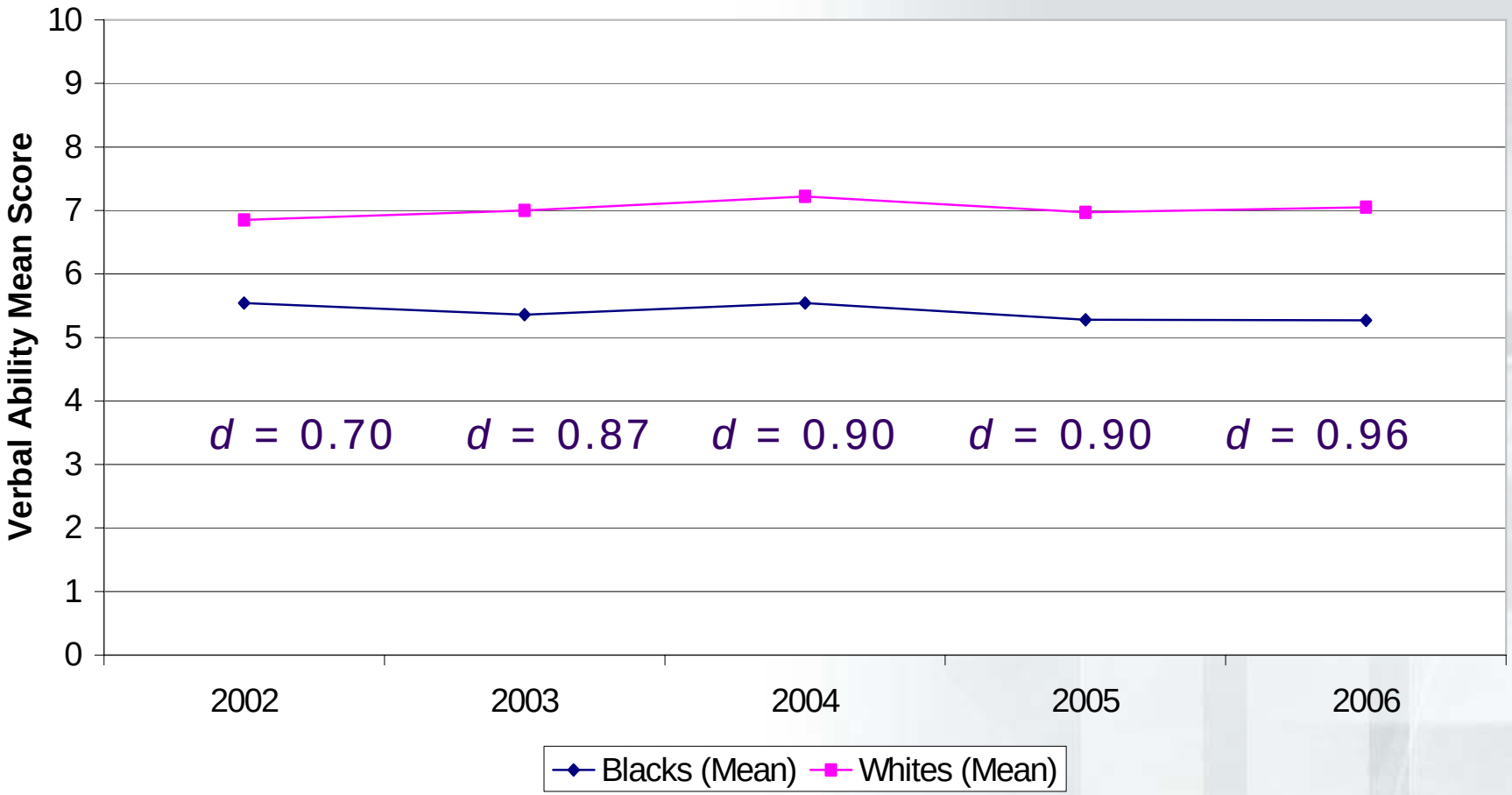
- International literature report group differences (Black/White) of on average 1 SD with cognitive ability.
  - > Hunter & Hunter (1984).
  - > Roth, BeVier, Bobko, Switzer & Tyler (2001).
- Examining SHL South Africa's database of completed ability assessments reflects similar patterns with White groups, on average, outperforming Black groups.

The influence of Internet administration on mean group differences

PEOPLE PERFORMANCE

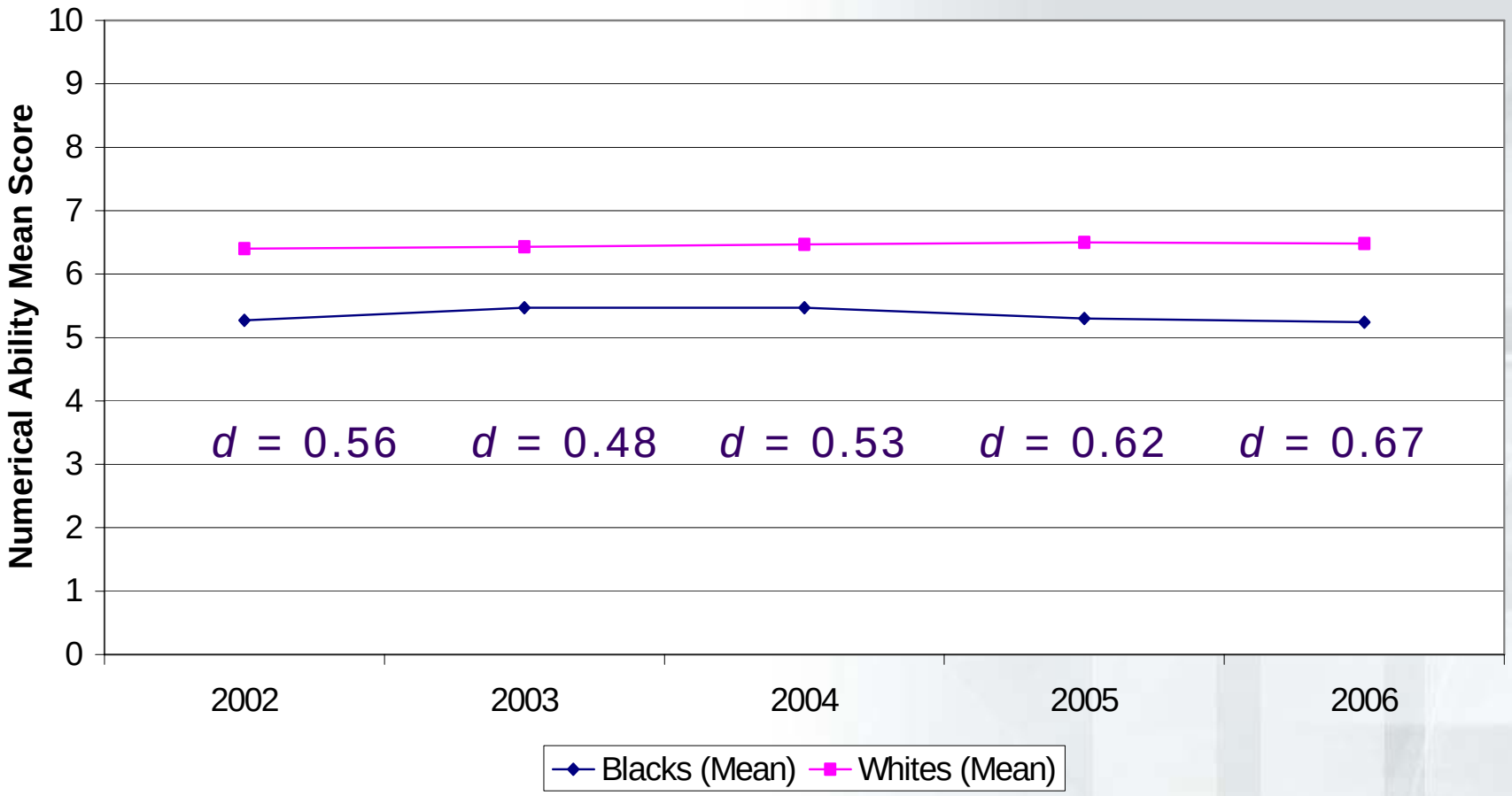
- Over five years SHL has gathered online (unsupervised) ability test data for clerical applicants to a financial institution (N=14,399):
  - > Aged between 16 and 63 (mean 25.25)
  - > 38% Male, 62% Female
  - > 86% Black, 13% White, 1% Other / Not Provided
- Research Question: Will internet administration change the effect size of Group differences between applicants? (Internet or computer skills may serve as a 'pre-screen' to this assessment).

Group differences for online verbal ability test over five years (2002-2006)



For the five years, *d* = 0.87

Group differences for online numerical ability test over five years (2002-2006)



For the five years, *d* = 0.55

The influence of Internet administration on mean group differences

PEOPLE PERFORMANCE

- The group differences online are similar, although somewhat smaller, than those observed in other (offline) data sets.
- However, similar (smaller) studies in online graduate recruitment do not support the observation of smaller differences between ethnic groups when completing ability tests online.
- We cannot make a conclusive statement yet, but the data of 14,000+ individuals suggests there is more to be explored in this regard.

Group Differences in South Africa: Personality Measures

- Internationally, differences between races on mean scores for personality measures are negligible:
  - > Bobko, Roth & Potosky (1999)
  - > Schmitt, Clause & Pulakos (1996)
- Research conducted by SHL South Africa with the OPQ32 mirrors this trend (SHL, 2006):
  - > Very small to moderate mean differences observed
  - > Internal consistency shown for both ethnic groups

Culture Fairness & the OPQ32i (SHL, 2006)

- Sample drawn from SHL's OPQ32i database (N=20,132 individuals across industries):
  - > 43% Male, 57% Female
  - > 71% Blacks, 29% Whites
  - > Grade 12 (42%), Degrees (29%) & Postgraduate (14%)

- The study's approach was three-fold:
  - > Comparing mean scores of the ethnic groups by calculating effect sizes (*d* statistic)
  - > Exploring the OPQ32i's internal consistency for the groups
  - > Comparing the covariance structures of the samples using Structural Equation Modelling

Standardised mean differences between Blacks & Whites on the OPQ32i

- Only very small to moderate differences found in the mean scores for the Black and White groups.
- The highest *d* statistic is a moderate effect size *d* = 0.48 on Decisive, and *d* = 0.40 on Controlling and Forward Thinking.
  - > 7 out of 32 Scales *d* > 0.30
  - > 15 out of 32 Scales *d* > 0.10 but < 0.30
  - > 10 out of 32 Scales *d* < 0.10

Internal Consistency of the OPQ32i for Black and White Groups

PEOPLE PERFORMANCE

- The internal consistency of the OPQ32i, for the two ethnic groups, was calculated using Cronbach's Alpha.

- For the Black group (N=14,367):
  - > Alphas range between 0.57 and 0.84
  - > Average reliability of 0.69

- For the White group (N=5,765):
  - > Alphas range between 0.65 and 0.87
  - > Average reliability of 0.78

Equivalence of Covariance Structures for Black and White Groups

PEOPLE PERFORMANCE

- A comparison of covariance structures was carried out using Structural Equation Modelling with EQS.
- The fit was tested of a model that evaluated each pair of intercorrelations, and each pair of scale variances, to be equal across the samples – Black and White.
- The *comparative fit index* (CFI; Bentler, 1990)
  - > **CFI** was **0.960**, indicating exceptionally good fit
- The *root mean square error of approximation* (RMSEA)
  - > **RMSEA** was **0.026**, indicating exceptionally good fit

Equivalence of Covariance Structures for Black and White Groups

- The analysis suggests that the pattern of relationships between scales can be considered the same for Black, Coloured, Indian and White ethnic groups as well as the language groups.
- Analysis of covariance structures for the sets of data suggest that these can be considered to be samples from the same population of scale relationships.
- The evidence supports the use of the OPQ32i in South Africa for Black and White population groups.

What does this mean?

- We know that group differences on cognitive measures exist between Black and White groups, with White groups typically scoring higher.
- This means that Adverse Impact could be a problem, as Black groups typically score lower.
- In line with international trends, group differences between Blacks and Whites on personality measures are minimal.

The Predictive Validity of Assessments for Blacks vs. Whites

PEOPLE PERFORMANCE

The Predictive Validity of Assessments for Blacks vs. Whites

PEOPLE PERFORMANCE

- An objective assessment “myth” is that assessments cannot be used to predict job performance for Black candidates.
- Validation data over the years in South Africa reveals a trend that the predictive validities of assessments tend to be higher for Black than for White groups.
- Three studies selected from SHL’s validation databank illustrate this:
  - > Technical Officers in Telecommunications (1999)
  - > Business Advisory Services in Finance (2005)
  - > Call Centre Operators in Insurance (2006)

Using the TTB for Selection in the Telecommunications Industry (1999)

The TTB in a Telecommunications Organisation (1999)

PEOPLE PERFORMANCE

- Concurrent validation study to assess the effectiveness of using the Technical Test Battery (TTB) in selecting Technical Officers in a Telecommunications organisation.

- Sample consisted of 107 Technical Officers:

  > Aged between 19 & 49 (Mean 32.38)
  > Majority Male (99%)
  > 61% Black, 32% White, 7% Not Provided
  > Majority (49%) held Grade 12

The TTB in a Telecommunications Organisation (1999)

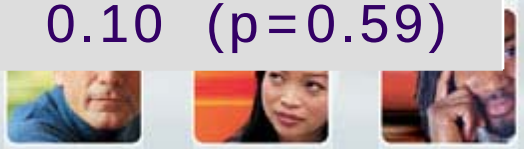
- Group differences of small to medium effect size observed between 0.08 & 0.6 SD, with the White Group typically scoring higher:

Measure	*d*	Black (N)	White (N)
Verbal Comprehension	0.33	66	34
Visual Estimation	0.60	66	34
Technical Understanding	0.21	66	34
Numerical Reasoning	0.08	66	34
Fault Diagnosis	0.10	61	29

The TTB in a Telecommunications Organisation (1999)

- Between predictors and the criterion measure, larger correlations are observed for the Black group (except for Visual Estimation):

Predictor	Correlation Coefficient	
	Blacks	Whites
Verbal Comprehension	**0.47 (p=0.00)**	0.32 (p=0.07)
Visual Estimation	0.10 (p=0.12)	**0.29 (p=0.10)**
Technical Understanding	**0.42 (p=0.00)**	0.34 (p=0.05)
Numerical Reasoning	**0.43 (p=0.00)**	0.31 (p=0.07)
Fault Diagnosis	**0.51 (p=0.00)**	0.10 (p=0.59)



The TTB in a Telecommunications Organisation (1999)

PEOPLE PERFORMANCE

	Multiple R	Multiple R^2
1st Step: > Technical Officers' Test Battery	0.48	0.23
2nd Step: > Race entered into the equation	0.49	0.24

- Race does not explain any additional statistically significant variance (p=0.200)

  > Intercepts & slopes of Black & White groups similar
  > No evidence of unfairness (according to Cleary's Model)
  > A common regression line can be used to predict job success

PEOPLE PERFORMANCE

Using the CRTB for Staff Selection in a Financial Institution (2005)

The CRTB in a Financial Institution (2005)

- The study determined the effectiveness of the CRTB as a selection tool for various positions at a financial institution.

- Sample consisted of Various Business Advisory Services (N=165)

  > Aged between 20 & 46 (Mean 29.26)
  > 43% Male, 57% Female
  > 45% Black, 55% White
  > Majority Degree (19%) or Postgraduate (52%)



The CRTB in a Financial Institution (2005)

- Group differences in the region of 1 SD observed on the ability tests, with the White group scoring on average higher:

Measure	*d*	Blacks (N)	Whites (N)
Verbal Reasoning	0.99	47	57
Numerical Reasoning	1.03	47	56
Diagrammatic Reasoning	1.14	44	52
Total Ability	1.26	48	57
Job Performance Rating	0.77	48	57



The CRTB in a Financial Institution (2005)

- Between predictors and the criterion measure, larger correlations are observed for the Black group:

Predictor	Correlation Coefficient	
	Blacks	Whites
Verbal	**0.42 (p=0.00)**	0.16 (p=0.24)
Numerical	**0.36 (p=0.01)**	0.22 (p=0.11)
Diagrammatic	**0.35 (p=0.02)**	0.32 (p=0.02)
Total Ability	**0.45 (p=0.00)**	0.36 (p=0.00)

The CRTB in a Financial Institution (2005)

	Multiple R	Multiple R^2
1st Step: > Three CRTB Ability Tests	0.49	0.24
2nd Step: > Race entered into the equation	0.51	0.26

- Race does not explain any additional statistically significant variance (p=0.130)
  - > Intercepts & slopes of Black & White groups similar
  - > No evidence of unfairness (according to Cleary's Model)
  - > A common regression line can be used to predict job success

Predicting Call Centre Operator Performance in Insurance (2006)

PEOPLE PERFORMANCE

Predicting Call Centre Operator Performance in Insurance (2006)

PEOPLE PERFORMANCE

- Concurrent validation study to assess the effectiveness of using personality assessment and Person-Job Matching (PJM) to predict the job performance of call centre operators in the insurance industry.

- Sample consisted of 233 Call Centre Operators:

  > Aged between 19 & 58 (Mean 27.93)
  > 45% Male, 55% Female
  > 75% Black, 25% White
  > Grade 12 to a higher degree

Predicting Call Centre Operator Performance in Insurance (2006)

- The Black group shows higher Person-Job Match scores, while the White group shows higher criterion performance:

Measure	*d*	Blacks (N)	Whites (N)
Person-Job Match Score	-0.29	175	58
Criterion (Incentive Bonus)	0.24	175	58

Predicting Call Centre Operator Performance in Insurance (2006)

- A larger correlation is observed between Person-Job Match scores for the Black Group than for the White Group:

Predictor	Correlation Coefficient	
	Blacks	Whites
Person-Job Match Score	**0.38 (p=0.00)**	0.27 (p=0.04)

Predicting Call Centre Operator Performance in Insurance (2006)

Predicting Job Success	Multiple R	Multiple R^2
1st Step: > Call Centre Operator PJM Scores	0.33	0.11
2nd Step: > Race entered into the equation	0.36	0.13

- Entering Race makes a statistically significant contribution to the regression equation (p=0.017)
- 3rd Step: Entered an interaction term (PJM Score x Race) to test for slope and intercept differences between the two race groups

Predicting Call Centre Operator Performance in Insurance (2006)

Predicting Job Success	Multiple R	Multiple R^2
1st Step: > Call Centre Operator PJM Scores	0.33	0.11
2nd Step: > Race entered into the equation	0.36	0.13
3rd Step: > Interaction term entered	0.36	0.13

- Entering the interaction term makes no statistically significant contribution (p=0.630), suggesting that:
  - > While there are intercept differences between the groups
  - > There are no significant differences between the slopes

Predicting Call Centre Operator Performance in Insurance (2006)

PEOPLE PERFORMANCE

- In terms of Cleary's Model of Test Fairness:
  - > The White group's intercept lies above the Black group
  - > However, the slopes of Black & White groups are similar
  - > Using a common regression line will under-predict success for Whites and over-predict success for Blacks
  - > Permissible according to South Africa's employment equity legislation
  - > Such action would not constitute unfair discrimination against the Black group

What does this mean?

- These studies suggest that:
  - > Job performance can be predicted more accurately for Black groups (i.e. Differential Validity).
  - > In most cases, predicted job performance does not differ based on an individual's group membership (i.e. Differential Prediction).

Remember:

Differential Validity ≠ Differential Prediction

PEOPLE PERFORMANCE

Changing Demographics of Applicants: What's happening in practice?

Changing Demographics of in Validation: 1994 - 2007

- Initial validation studies conducted by SHL South Africa reported sample racial distributions as being typically majority White participants, or roughly equal distributions (where such data was available):

Study	Black	White	Missing
ATC Cadets (1997; N=75)	39%	61%	-
Middle Managers (1999; N=72)	28%	72%	-
Demand Forecasters (2000; N=90)	60%	40%	-



Changing Demographics of in Validation: 1994 - 2007

- More recent validation studies conducted by SHL South Africa report increasingly majority Black participants...

Study	Black	White	Missing
Business Advisory Services (2005; N=165)	74 (45%)	90 (55%)	1 (0.6%)
Call Centre Agents (2006; N=107)	82 (77%)	7 (7%)	18 (17%)
Yard Officials (2007; N=107)	104 (97%)	3 (3%)	-



Graduate Recruitment Data: 2004 - 2007

PEOPLE PERFORMANCE

- What impact do these changing demographics have on selection decisions?
- To answer this question, we examined data for graduate recruitment, gathered over four years.
- The process followed for graduate recruitment entails:
  - > Online application and CV screening (GNDs)
  - > Online competency-based behavioural questionnaires
  - > Online completion of ability & personality measures
  - > From here on the organisation applies PDI quotas

Graduate Recruitment Data: 2004 - 2007

- The percentage Black applicants increases over the years:

Year	Black	White	Other
2004	2709 (80%)	559 (17%)	116 (3%)
2005	2516 (87%)	352 (12%)	24 (1%)
2006	1007 (85%)	160 (14%)	18 (1%)
2007	2112 (91%)	189 (8%)	24 (1%)

Graduate Recruitment Data: 2004 - 2007

- Group differences observed in the group across the 4 years range from 0.5 – 1 SD in favour of White group:

Year	d-statistic (Total Group)		
	Verbal Reasoning	Numerical Reasoning	Combined Ability
2004	1.03	0.95	1.21
2005	1.10	0.87	1.23
2006	0.50	0.71	0.72
2007	0.97	0.73	1.01

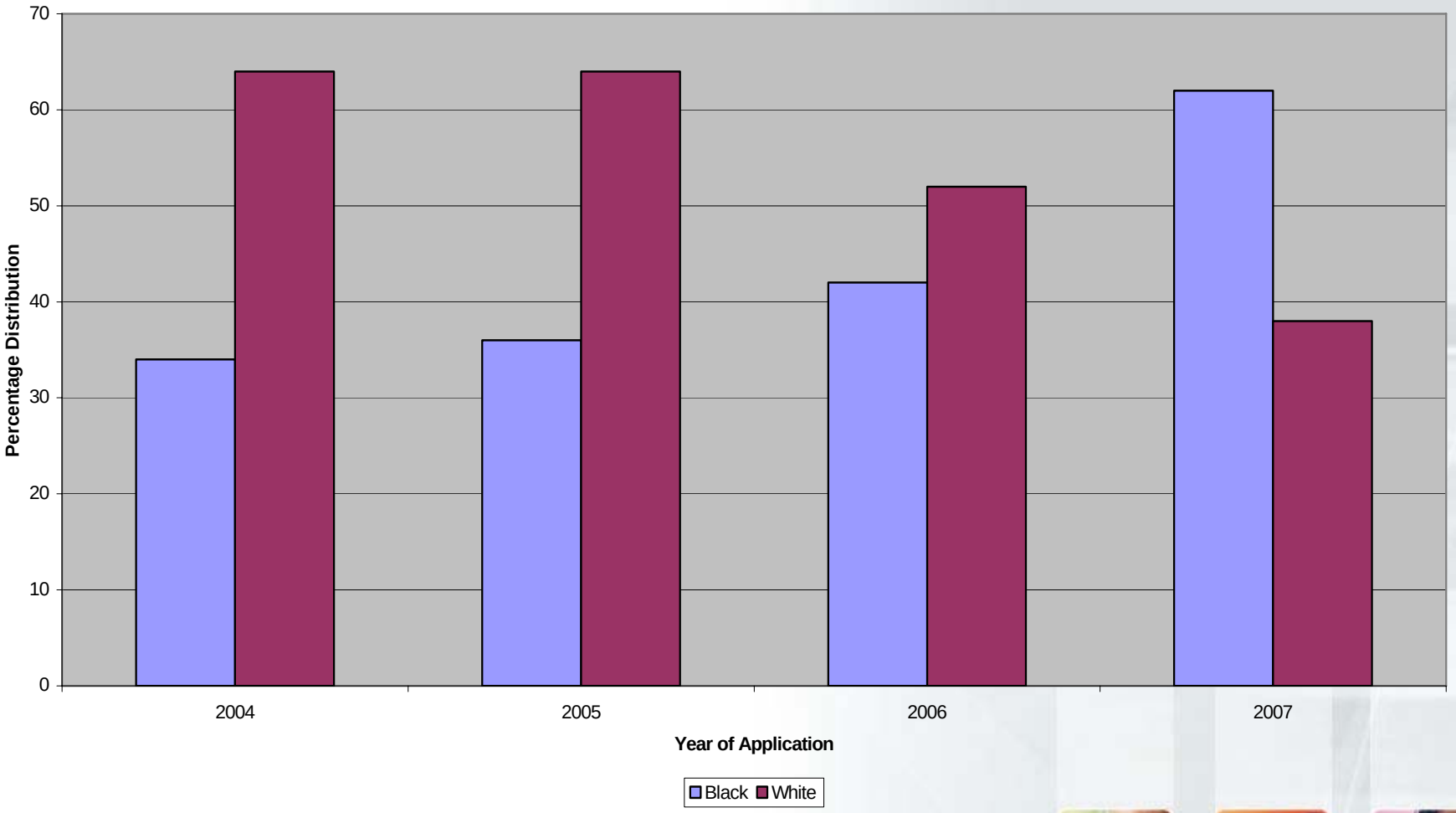
Graduate Recruitment Data: 2004 - 2007

- However, if you look at the racial distribution for the top 50 candidates (based on ability):

Top 50 Candidates Based on Ability			
Year	**Black**	**White**	**Other**
2004	17 (34%)	32 (64%)	1 (2%)
2005	18 (36%)	32 (64%)	-
2006	21 (42%)	26 (52%)	3 (6%)
2007	31 (62%)	19 (38%)	-



Racial Distribution: The Top 50 Ability Rank List



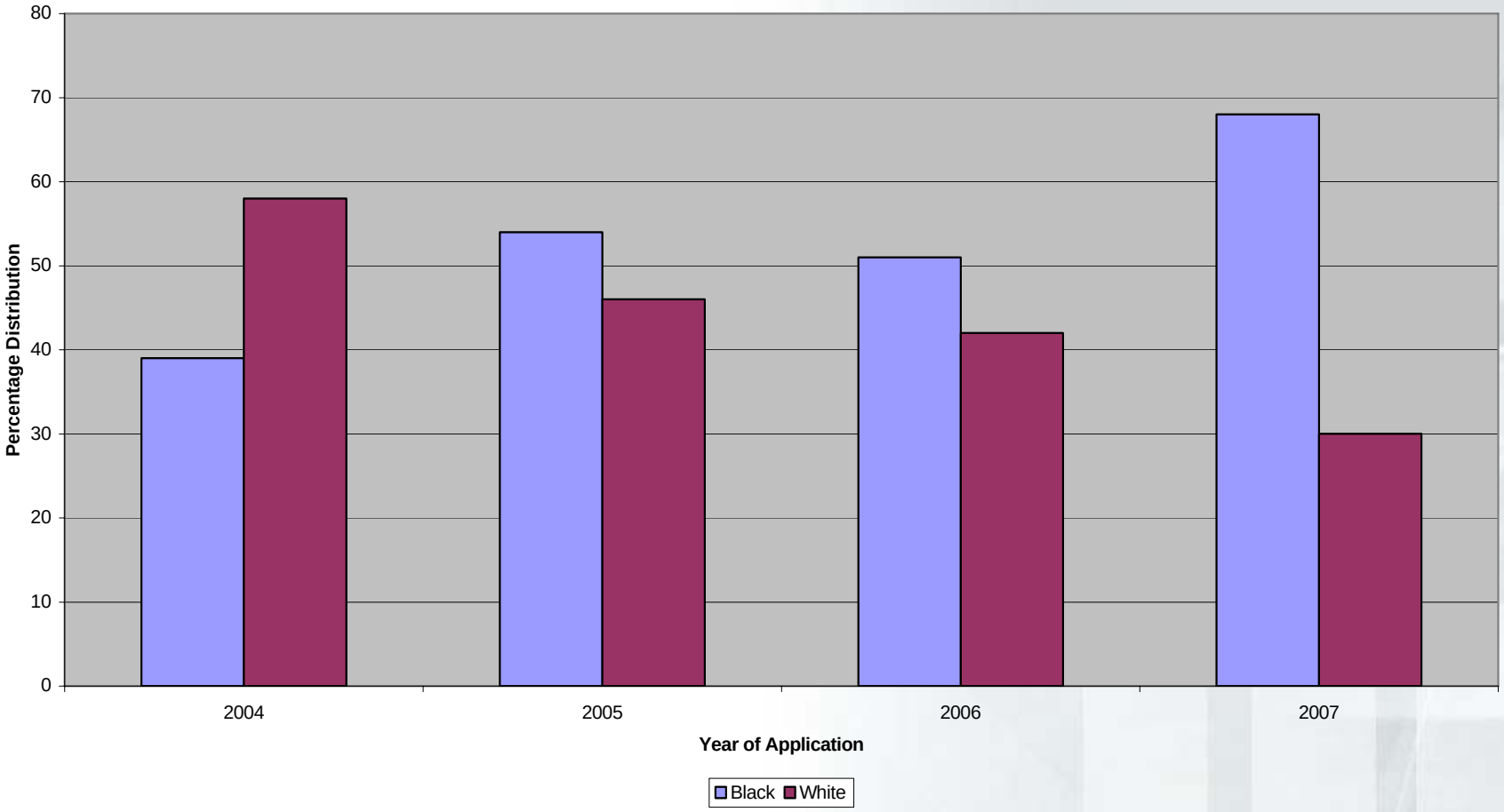
Graduate Recruitment Data: 2004 - 2007

- This trend continues if you look at the racial distribution for the top 100 candidates (based on ability):

Top 100 Candidates Based on Ability			
Year	**Black**	**White**	**Other**
2004	39 (39%)	58 (58%)	3 (3%)
2005	54 (54%)	46 (46%)	-
2006	51 (51%)	42 (42%)	7 (7%)
2007	68 (68%)	30 (30%)	2 (2%)



Racial Distribution: The Top 100 Ability Rank List





The Effect of Combining Personality Data with the Ability Data

- Equal Opportunities literature highlights how including assessments less likely to cause adverse impact can minimise the impact of group differences observed on cognitive ability tests.

- Rank listing the graduate applicants based on combined personality and ability data (Person-Job Matching) supports this.

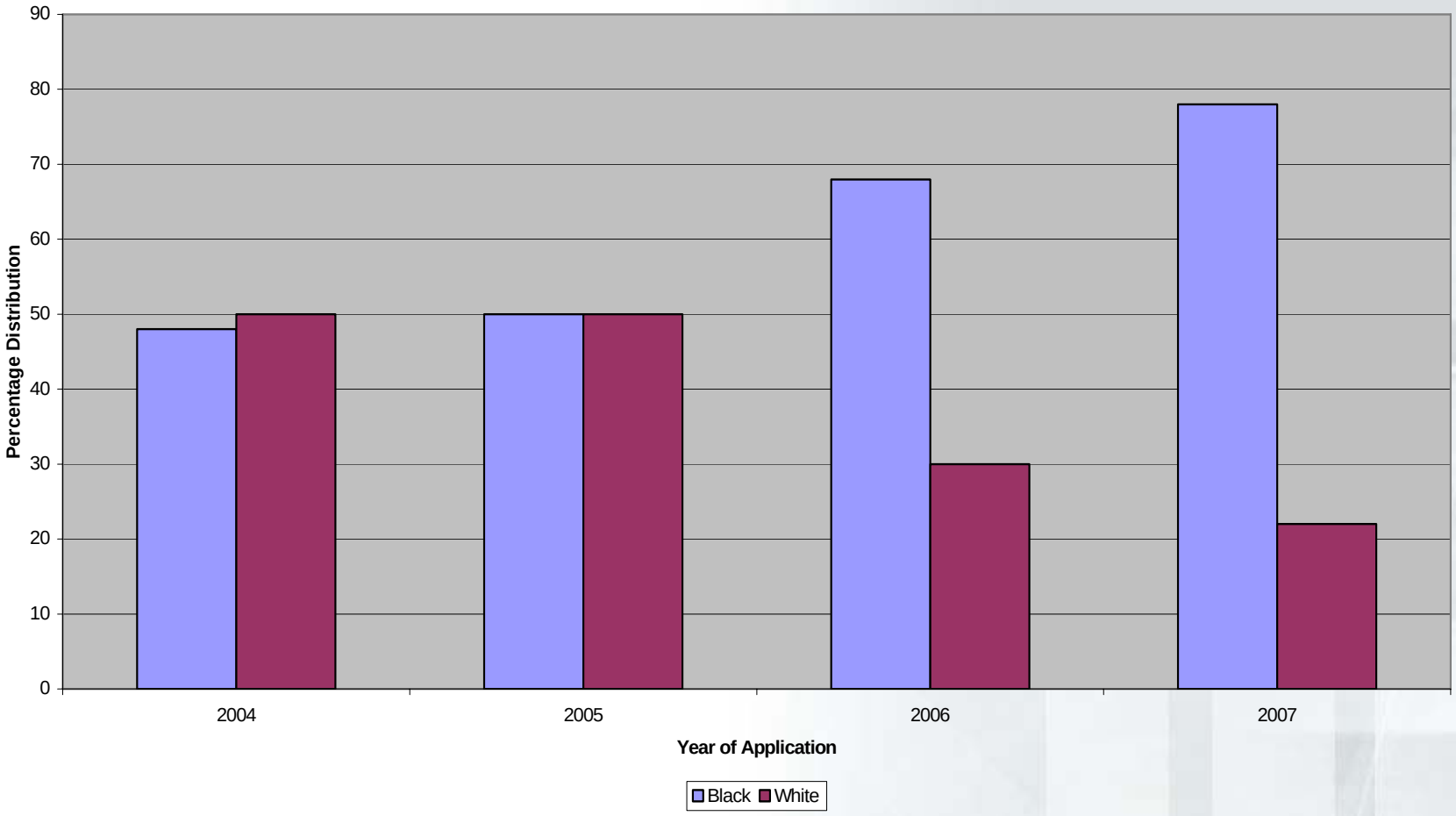
Top 50 Applicants Based on Person-Job Matching (PJM)

- The racial distribution of the top 50 applicants has changed from roughly equal (in 2004) to almost 80% Black (in 2007):

Top 50 Candidates Based on PJM			
Year	**Black**	**White**	**Other**
2004	24 (48%)	25 (50%)	1 (2%)
2005	25 (50%)	25 (50%)	-
2006	34 (68%)	15 (30%)	1 (2%)
2007	39 (78%)	11 (22%)	-



Racial Distribution: The Top 50 Merit List (Combined Personality & Ability Results)





In Conclusion

In Conclusion

- We know that group differences exist between Black and White groups on cognitive ability measures (less so for personality measures).
- We see that assessments can predict job performance more accurately for Blacks than for Whites.
- We find that, as the demographics of applicant pools shift increasingly more in favour of Black individuals, the impact of group differences – when using a top-down selection approach – is diluted.

In Conclusion

- In order to balance economic needs and social needs, some final thoughts to ponder:
  - > Perhaps too much emphasis has been placed on group differences at the cost of how assessment results can be used in fair decision-making processes.
  - > Need to consider the total picture of: Validity, applicant pools, selection cut scores, legislative context