Available online at www.sciencedirect.com
ScienceDirect
Learning and Individual Differences 18 (2008) 144–151
Learning and Individual Differences
www.elsevier.com/locate/lindif

Ethnic, gender, and socio-economic group differences in academic performance and secondary school selection: A longitudinal analysis ☆

Norah Frederickson [a,*], K.V. Petrides [b]

[a] Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK
[b] Institute of Education, University of London, UK

Received 12 February 2004; accepted 30 September 2005

Abstract

This study examined gender, socio-economic (SES), and ethnic group differences in academic performance (measured at 14 and 16 years) in a sample of 517 British pupils (mean age=16.5 years). White pupils outperformed their Black and Pakistani counterparts and high SES pupils consistently outperformed their low SES counterparts. Results from two Multiple Indicators Multiple Causes (MIMIC) models showed that controlling for IQ variance minimizes these group differences. The MIMIC models also revealed that Pakistani pupils and girls tend to underperform academically relative to White pupils and boys, respectively, at 14 years, once IQ and SES have been partialed out. These and other, more specific, findings are discussed with reference to predictive test bias, selection and streaming procedures, and implications for educational policy.

Keywords: IQ; Attainment; Ethnic; Gender; Socio-economic; School selection

Since the early 1980s there has been concern in Britain about the low educational achievement of particular minority ethnic groups and consequent exclusion from higher education and employment opportunities important to social participation and advancement (Rampton, 1981; Social Exclusion Unit, 1999; Swann, 1985). However, research on ethnic group differences in academic performance has produced conflicting conclusions. In part, this is attributable to confounds with, and different means of controlling for, socio-economic background. Ethnic differences may be over estimated, if account is not taken of socio-economic differences between different ethnic groups (Sammons, 1995). What also appears to be of relevance is the geographical location and time period in which studies have been carried out. Gillborn and Mirza (2000) report a national picture in which performance in the General Certificate of Secondary Education GCSE examinations at age 16 years by Black, Bangladeshi and Pakistani pupils is lower than that of White UK pupils. However, they also identify areas of the UK, where each of these ethnic groups performed better than any other at GCSE. This finding cautions against any stereotypical focus on supposed group characteristics and supports other research internationally that suggests a need for interactional analyses, incorporating features of the educational and broader environmental context (Portes, 1999).

Analyses of changes over time have also presented a variable picture. Cross sectional data from the Youth Cohort Studies (DfES, 2003) indicate that, although the performance of all ethnic groups at GCSE increased between 1991 and 1999, the only minority ethnic groups to make substantially greater gains than the White UK group were the Indian and 'other' Asian groups. In these two groups, equal or higher proportions of students achieved the top GCSE grades (A to C) compared to the White UK group, in at least five or more GCSE examination subjects. Looking at improvements over the years 1991–1999, two groups (Black and Bangladeshi) recorded similar, or marginally better, percentage improvements than the White UK group. However, the Pakistani group made very little improvement, with only 3% more students in 1999, as compared to 1991, achieving more than five GCSE results at grades A to C. This contrasts with the 13% increase for White students, 16% for Black students, 15% for Bangladeshi

☆ This study is part of a research programme supported by Nuffield Foundation grant SGS/00777/G.
* Corresponding author.
E-mail address: n.frederickson@ucl.ac.uk (N. Frederickson).

1041-6080/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.lindif.2005.09.001

students, 22% for Indian students, and 26% for other Asian students. There is therefore little evidence that national inequalities in attainment have been diminishing through the 1990s.

Longitudinal studies conducted in the 1980s had suggested that inequalities in attainment reduce with age and years in education. Sammons (1995) reported that the GCSE performance in 1989 of African Caribbean students in the study sample of Inner London schools did not differ significantly from that of White UK students. The performance of White UK students was significantly poorer than that of students from Asian backgrounds and an 'other' group comprising students from Chinese, South East Asian, Turkish, Greek and mixed race backgrounds. The only group to perform significantly more poorly at GCSE than the White UK students was those who had been rated by their teachers as not fully fluent in English in Year 3. When relative progress was examined by Sammons (1995), controlling for reading performance in Year 6 and Verbal Reasoning Band estimates from primary school teachers, all three minority ethnic groups had made significantly greater gains than the White UK group. Similar findings were reported by Nuttall, Goldstein, Prosser and Rasbash (1989), again in Inner London. Nuttall et al. (1989) found that all minority ethnic groups except the African Caribbean group showed better GCSE performance in the years 1985–1987 than the White UK group, from which the African Caribbean group did not differ. Nuttall et al. (1989) acknowledge that their study was limited by its lack of socio-economic data, which was particularly problematic in this case because the White UK population of Inner London schools was not considered by the authors to be nationally representative either socially or economically. Sammons (1995) had taken account of socio-economic differences, but also recognised the need to examine whether the findings applied to other parts of the UK, and, in particular, to areas that were not targeting ethnic minority achievement to the same extent as was the Inner London Education Authority in the 1980s.

A need to consider gender effects in the attainment of different ethnic groups was also identified (Sammons, 1995). At primary school, the overall performance of girls has long exceeded that of boys, although relative performance has been found to differ across different subject areas (Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988) and, more recently, within different ethnic groups (Demie, 2001). At GCSE, the proportion of pupils obtaining five or more grades A to C has shown an increasing 'gender gap', from 1989, when girls showed an advantage over boys of 3%, to 2002, when the advantage had increased to 10% (DfES, 2003). Gillborn and Mirza (2000) report that, since 1995, this gender gap has been apparent within each ethnic group, regardless of social class background. Longitudinal studies that have controlled for attainment at the end of primary schooling have found an increase with age through secondary school in disparities in attainment across gender and socio-economic status (Sammons, 1995).

More recent longitudinal analyses of ethnic group effects present a mixed picture. Richardson and Wood (1999) examined data from ten LEAs in or near London on the achievement of Black pupils on Key Stage 2 English assessments at age 14 and the attainment of five or more grades A to C passes in GCSE examinations at age 16. Compared to the national average, rather than to the attainment of other ethnic groups in the areas from which the Black sample was drawn, the performance of the Black group was found to be relatively worse at GCSE than at KS2. However, data provided by LEAs in their submissions to the DfES for support from the Ethnic Minority Achievement Grant in 1998, also indicated a decline in the relative examination performance of African Caribbean students between the ages of 14 and 16 years (Gillborn & Mirza, 2000).

Haque and Bell (2001) examined the academic performance in national Key Stage 3 (KS3) tests at 14 years and in GCSE at 16 years for pupils in 12 schools across a number of LEAs where there was a high proportion of Bangladeshi pupils. They found that performance on Key Stage 3 assessments for all the minority ethnic groups identified (African, Bangladeshi, Indian, Pakistani and Other) was significantly lower than the performance of the White UK group. The addition of a measure of social class to the analysis reduced, but did not eliminate, the effects of ethnicity. However, by 16 years, when GCSE examinations were taken, only the Indian group differed significantly from the White UK group, the difference favouring the former. An examination of progress during Key Stage 4 indicated that progress was least for the White UK group, with the Bangladeshi group next lowest.

It appears that the discrepancies between the cross sectional and the longitudinal studies may well relate to underlying differences in their samples. While cross sectional data has mainly come from large national samples, longitudinal studies have been predominately carried out in geographical areas where the proportions of students from minority ethnic groups is very high and the white students are more socially and economically disadvantaged than the national average. Where longitudinal studies have drawn comparisons with national averages (Richardson & Wood, 1999) or included areas with more nationally representative student populations (Gillborn & Mirza, 2000) their findings have tended to be similar to those drawn from cross sectional data. However, such tentative conclusions must be qualified by the paucity of relevant longitudinal data. In particular, the lack of longitudinal studies examining the progress of Pakistani pupils in areas with nationally representative proportions of these students constitutes a problematic gap in the literature, given that this group has shown the lowest percentage rise in GCSE achievement during the last decade. The first aim of the present study was to address this gap.

A longitudinal analysis of progress between the end of KS2 and GCSE assumes particular importance given reports of increasing use of formal and informal mechanisms for secondary school selection (Edwards & Tomlinson, 2002). If, as is claimed by Sammons (1995) and Haque and Bell (2001), relatively greater gains are made in secondary school by pupils from minority ethnic groups, then selection on the basis of performance at 11 years is likely to discriminate against them, as it will under predict their performance at GCSE. This is the second issue to be investigated in this paper. Our focus will be on the issue of secondary school selection because of the availability of consistent data across schools within a selective LEA. However, the issue is much wider. Since the 1988 Education Act introduced a National Curriculum in the UK, many schools have established practices of setting or streaming students on the basis of national Key Stage 2 (KS2) results or cognitive ability test data collected in the first term in

secondary school (Ireson & Hallam, 2001). With the introduction of 'tiered' GCSE examinations, the tier for which a pupil is entered can limit the result achieved, so, for example, a candidate entered for a Foundation Tier cannot achieve a C grade or higher. There are reports that students from lower streams or sets are being entered for lower tiers (Ireson & Hallam, 2001) and that Black pupils are disproportionately likely to be entered in the lowest tier (Gillborn & Youndell, 2000).

While criticisms of the secondary selection tests played a role in promoting the spread of comprehensive education in the UK in the 1960s, there has been little recent discussion of the selection procedures employed by those LEAs that did not introduce comprehensive schooling, retaining instead grammar and secondary modern schools (Crook, Power, & Whitty, 1999). As was the case 50 years ago, the central element of the grammar school selection procedures in use today are group cognitive ability tests. Vernon (1957) recommended that the tests be referred to as 'academic aptitude' tests, rather than 'cognitive ability' tests in order to avoid appearing to make the claim "that they are measuring purely innate ability, in contrast to acquired attainments" (p. 173). Had more attention been paid to this recommendation, the interpretation of well established score differences between ethnic groups on standardized cognitive ability tests may have caused less controversy (Laosa, 1996; Mackintosh, 1986). From an extensive review of the literature, Mackintosh (1998) concluded that the research evidence was consistent with the view that 'if environmental differences between blacks and whites could be miraculously eliminated, the two groups might well obtain approximately equivalent IQ scores' (p. 156). In common with other recent reviews, it is argued that while measured IQ may have a significant heritable component, there is little evidence that genetic explanations are required, even in part, to account for average differences between groups (Howe, 1997).

The use of IQ tests in educational decision making has been almost as controversial as the interpretation of between group differences. In making selection decisions for special education there has been extensive debate about ethnic group bias in IQ tests, particularly those that are verbally based (Hillard, 1994; Reschly & Ward, 1991). However there has been little specific discussion of this issue in relation to the use of IQ tests in secondary selection decisions. The second aim of this study was to investigate possible ethnic group bias in a commonly used secondary selection test. It is recognised that score differences between particular groups on a selection test do not necessarily indicate that the test is biased. A widely accepted definition states 'bias is differential validity of a given interpretation of a test score for any definable, relevant subgroup of test takers' (Cole & Moss, 1989). The primary purpose of the secondary selection process is to identify for grammar school places the pupils most likely to achieve academic success during the secondary phase of their education, where academic success is defined principally in terms of GCSE results at 16 years. The key question to be addressed, therefore, is whether the predictive validity of the secondary selection tests is the same across different ethnic groups. If the predictive validity is equivalent, then the tests are equally good (or equally poor) predictors of future performance for individual pupils from different ethnic groups. If the predictive validity is not equivalent across different ethnic groups, the selection tests will underestimate the subsequent achievement of some groups of pupils relative to others. As such, the tests could be said to be unfair to or biased against these groups of pupils.

In summary, this study sets out to examine academic attainment in secondary schools of students from different ethnic and gender groups in an LEA which uses verbal reasoning tests for secondary school selection. It was hypothesised that different groups would make different rates of progress, so that boys would progress less well than girls and Black and Pakistani pupils would progress less well than White UK pupils in KS3 at 14 years and in GCSE at 16 years. It was also hypothesised that the predictive validity of the verbal reasoning test scores would be similar across the different gender and ethnic groups. Given concerns about the extent to which ethnicity analyses tend to be confounded by socio-economic status (Laosa, 1996), this variable was also investigated.

1. Method

1.1. Participants

Questionnaire data were collected from 901 pupils in a selective LEA as part of a longitudinal survey into psychosocial influences on scholastic behaviour and achievement (Petrides, Frederickson, & Furnham, 2004). The composition by ethnic group of Year 11 pupils in the LEA was similar to that reported in the Youth Cohort Study conducted in the same year (2000): 1.7% Black pupils compared with 2.2% in the YCS sample, 4.8% Pakistani pupils compared with 2.1% in the YCS sample, 1.3% Indian pupils compared with 2.7% in the YCS sample. The pupils were drawn from urban, suburban, and rural areas of a large shire county. Thirty five percent of the pupils attended the two grammar schools that participated in the study and the rest attended one of the five secondary modern schools that took part.

Full data for the analyses that follow were available for 517 pupils. Approximately 53% of participants were males and 47% females. In the first wave of the study, all participants were Year 11 pupils in British secondary education (mean age of approximately 16.5 years). The three ethnic groups examined in this study were: Black (N=11, 2.1%; including 'Black African', 'Black Caribbean,' and 'Black Other'), Pakistani (N=29, 5.6%), and White UK (N=477, 92.3%). Due to the fact that there were only seven valid cases with full data, pupils of Indian origin were excluded from further analysis.

1.2. Measures

1.2.1. Ethnicity

Pupils were provided with the categories used for ethnic monitoring in the LEA that supported the study: Bangladeshi, Black African, Black Caribbean, Black Other, Chinese, Indian, Pakistani, White European, White-UK, White-Other and Other. These are the ethnic group names most commonly used in official statistics in the UK and in relevant academic research (Gillborn & Mirza, 2000). Due to the small number of pupils with minority ethnic backgrounds, certain groups were combined.

Thus, following the Youth Cohort Study, a 'Black' category was created, which combined the three Black groups.

1.2.2. Socio-economic status (SES)

Pupils were asked to indicate the occupations of their father and mother. These were subsequently coded on the six point scale used in the Youth Cohort Study prior to 2000: Managerial/ Professional, Other nonmanual, Skilled manual, Semi-skilled manual, Unskilled manual, Other/not classified. Pupils with at least one parent in a professional or skilled manual occupation were coded as 'high SES,' with the rest coded as 'low SES'.

1.2.3. Verbal Reasoning Test (VRT)

This tailor made test, developed by the National Foundation for Educational Research, measures primarily crystallized cognitive ability and is used by the educational authority that supported this study in secondary school selection. The internal consistency reliability of scores on this test (KR20) is usually in the order of about .97 (National Foundation for Educational Research, personal communication). The test is administered three times to each pupil and the score that the educational authority uses, and was made available to us, represents the average of the best two performances.

1.2.4. Key Stage 3 assessment (KS3) result

In the UK, pupils are statutorily assessed at the end of each of the four stages of the National Curriculum, which is followed by all publicly funded schools. Pupils will normally be about 14 years old when national testing occurs at the end of Key Stage 3. At this stage, attainment in the three core National Curriculum subjects of English, maths, and science is assessed and a numerical level assigned, with higher levels indicating better performance.

1.2.5. General Certificate of Secondary Education (GCSE) results

GCSEs are the principal means of assessing pupil attainment at the end of compulsory secondary education at 16 years. Assessment of GCSEs is usually by external examination and coursework, with the balance towards the former. Assessed subjects include English, math, science, religion, arts, music, design and technology, etc. Some of these are compulsory, whereas others are optional. GCSEs are graded from A* to G. Following Sammons (1995), these letter grades were assigned ascending numbers in sequence, starting from '1' that was assigned to the lowest pass grade.

1.2.6. Procedure

Schools were invited to participate by letter from the Director of Education. Phase 1 of the study involved completion by pupils in the Spring 2000 of a questionnaire battery, which was administered by teachers in class according to a detailed protocol. The questionnaire battery began with a letter to the pupils providing information about the study, offering assurances about confidentiality, and requesting their participation. The first section collected information on ethnicity, any languages spoken at home other than English, and parental occupation.

In Phase 2 of the study, which was carried out during Summer 2000 the following information was collected from the schools

Table 1
Means, SDs, and significance tests for VRT, KS3, and GCSE, broken down by gender, SES, and ethnicity

	Males	Females	*t*-test
VRT	112.35 (15.22)	111.43 (15.42)	.79
KS3 En	5.53 (1.40)	5.59 (1.12)	.66
KS3 Ma	5.61 (1.59)	5.28 (1.51)	2.95** (.21)
KS3 Sc	5.33 (1.41)	4.89 (1.50)	4.31** (.30)
GCSE En	4.89 (1.48)	5.42 (1.49)	5.05** (−.35)
GCSE Ma	4.77 (1.85)	4.66 (2.03)	.80
GCSE Sc	4.93 (1.82)	4.81 (1.88)	.92

	High SES	Low SES	*t*-test
VRT	114.18 (14.69)	102.96 (13.5)	6.66** (.79)
KS3 En	5.71 (1.24)	5.10 (1.21)	4.54** (.50)
KS3 Ma	5.70 (1.48)	4.62 (1.41)	6.72** (.75)
KS3 Sc	5.35 (1.40)	4.28 (1.39)	7.11** (.78)
GCSE En	5.37 (1.44)	4.50 (1.43)	5.71** (.61)
GCSE Ma	4.98 (1.87)	3.89 (1.86)	5.50** (.58)
GCSE Sc	5.11 (1.81)	3.88 (1.62)	6.56** (.72)

	White (a)	Pakistani (b)	Black (c)	*F*	Tukey tests
VRT	113.31 (14.67)	96.24 (12.99)	101.94 (14.31)	36.04** (.11)	a>b, a>c
KS3 En	5.64 (1.24)	4.73 (1.28)	5.09 (1.31)	18.38** (.05)	a>b
KS3 Ma	5.65 (1.50)	3.91 (1.21)	4.36 (1.47)	49.70** (.13)	a>b, a>c
KS3 Sc	5.30 (1.40)	3.60 (1.20)	4.57 (1.25)	49.69** (.13)	a>b, a>c, c>b
GCSE En	5.31 (1.48)	3.92 (1.26)	4.47 (1.31)	32.82** (.08)	a>b, a>c
GCSE Ma	4.90 (1.89)	3.18 (1.78)	3.71 (1.60)	30.96** (.08)	a>b, a>c
GCSE Sc	5.08 (1.80)	3.38 (1.47)	3.75 (1.45)	35.50** (.09)	a>b, c>b

Note: Abbreviations as in Fig. 1. Numbers in parentheses in *t*-test and *F* columns are effect sizes (Cohen's *d* and η^2, respectively). *$p<.05$, **$p<.01$.
Key: VRT = Verbal Reasoning Test, SES = Socio-economic Status, BL = Black, PA = Pakistani, KS3=Key Stage 3, GCSE = General Certificate of Secondary Education, En = English, Ma = Maths, Sc = Science (at KS3 level), gc_En = English, gc_Ma = Maths, gc_Sc = Science (at GCSE level).

Table 2
Intercorrelations of the variables in the study

	VRT	En	Ma	Sc	gc_En	gc_Ma	gc_Sc	SES	Sex
VRT	–	.678	.834	.787	.753	.818	.758	.380	−.022[a]
En	.	–	.649	.636	.728	.666	.654	.194	.034[a]
Ma	.	.	–	.804	.692	.839	.776	.271	−.099[b]
Sc	.	.	.	–	.669	.769	.814	.297	−.153
gc_En	.	.	.	.	–	.774	.735	.224	.212
gc_Ma	.	.	.	.	.	–	.833	.239	−.022[a]
gc_Sc	.	.	.	.	.	.	–	.267	−.040[a]
SES	.	.	.	.	.	.	.	–	−.067[a]

Note. N=717. All values significant at $p<.01$, except [a]p=ns, [b]$p<.05$.
Key: VRT = Verbal Reasoning Test, SES = Socio-economic Status, BL = Black, PA = Pakistani, KS3 = Key Stage 3, GCSE = General Certificate of Secondary Education, En = English, Ma = Maths, Sc = Science (at KS3 level), gc_En = English, gc_Ma = Maths, gc_Sc = Science (at GCSE level).

for the Year 11 pupils: KS3 results in English, maths and science (collected when the pupil was 14 years old) and the overall score obtained on the Verbal Reasoning Test (VRT) administered in the year before secondary transfer, when pupils were 11 to 12 years old. During the final phase of the study in Autumn 2000, schools provided GCSE results for Year 11 pupils.

2. Results

The means and standard deviations for the main variables of interest, broken down across gender, ethnicity, and SES, are presented in Table 1, along with tests of statistical significance. As can be seen in Table 1, boys outperformed girls in KS3 maths and KS3 science, whereas girls outperformed boys in GCSE English. There were extensive differences between high and low SES pupils, with the former outperforming the latter in every case. Last, the ANOVA comparing the three ethnic groups in the study showed that the White UK group consistently outperformed the Pakistani and Black groups. It should be noted that these analyses do not take into account the correlations between the variables involved. For example, many of the differences in the academic performance measures may disappear once VRT and SES scores are considered. The analyses below are designed to take all relevant interdependencies into account, thus providing a clear picture of the various group differences.

Two Multiple Indicators Multiple Causes (MIMIC; Jöreskog & Goldberger, 1975; Muthén, 1989) models were tested via Mplus 2.13 (Muthén & Muthén, 2001). This data analytic technique is especially suitable in this case, because of the small sample sizes of the minority groups. In all cases, the two latent variables of interest were KS3 (operationalized via three distinct indicators of academic performance at 14 years: English, math, and science) and GCSE (operationalized via three distinct indicators of academic performance at 16 years: GCSE English, GCSE math, and GCSE science). The intercorrelations of the variables in the study are given in Table 2.

MIMIC modelling is commonly used for the investigation of group differences on latent variables. Particularly interesting is Muthén's (1989) extension of the MIMIC technique to allow the regression of the indicators (i.e., the observed variables) directly on the exogenous variables. In the present case, this allows us to examine whether the groups of interest (e.g., males and females)

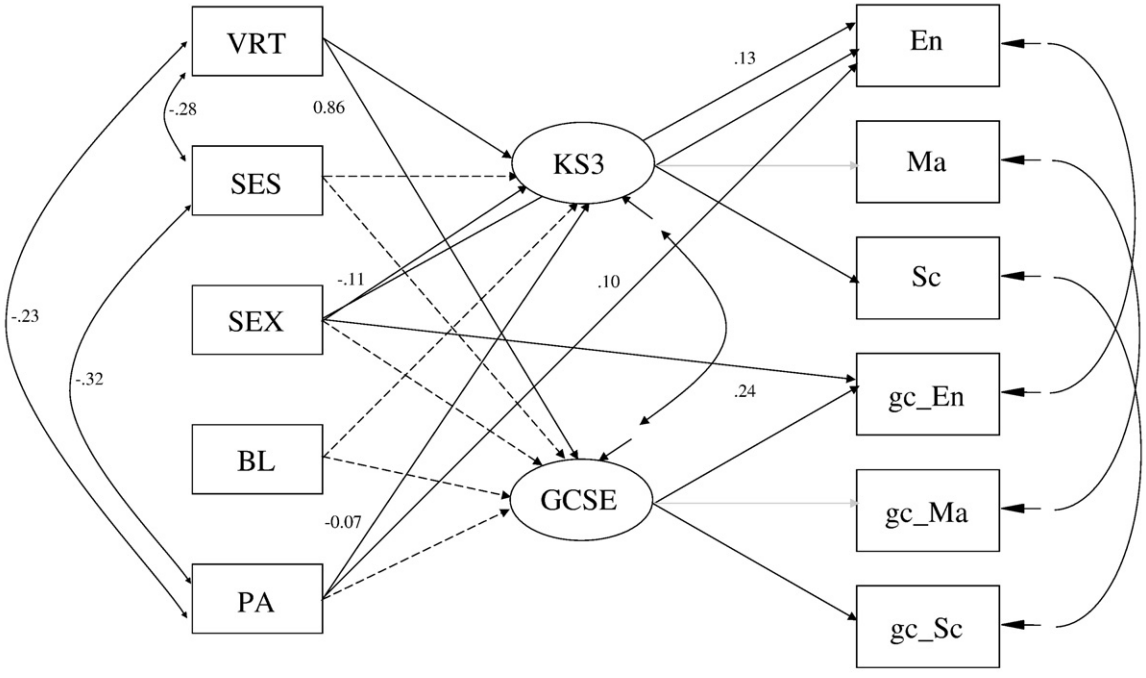

Fig. 1. Multiple Indicators Multiple Causes (MIMIC) model with completely standardized maximum-likelihood parameter estimates. Dashed lines denote paths that did not reach statistical significance. Solid grey lines denote the reference indicators for the two latent variables. To facilitate presentation, parameter estimates for the measurement part of the model are not depicted in the figure. VRT = Verbal Reasoning Test, SES = Socio-economic Status, BL = Black, PA = Pakistani, KS3 = Key Stage 3, GCSE = General Certificate of Secondary Education, En = English, Ma = Maths, Sc = Science (at KS3 level), gc_En = English, gc_Ma = Maths, gc_Sc = Science (at GCSE level).

differ on a particular indicator after controlling for any differences on the factor to which the indicator belongs (for example, the direct path from the Pakistani dummy variable on English in Fig. 1).

2.1. Model 1

A MIMIC model was set up, whereby KS3 and GCSE were modelled as latent variables and two dummy variables were introduced to represent the three ethnic groups of interest, with White UK pupils as the reference group (Fig. 1). In addition, SES, VRT, and sex were included as background variables in the model. The purpose of this analysis was to determine whether there are ethnic group differences in the common reliable variance of the six academic performance indicators (English, math, and science scores at KS3 and at GCSE level), *after controlling* for the aforementioned background variables.

The model provided a good fit to the data: $\chi^2_{(25)}=128.39$, $p<.01$, CFI=.97, SRMR=.03, RMSEA=.09 (for conservative cut off values, see Hu & Bentler, 1999). The completely standardized maximum-likelihood parameter estimates of the correlations between the exogenous variables and their effects on KS3 and GCSE are given in Fig. 1 for significant paths only. As can be seen from the correlations between the exogenous variables, Pakistani pupils and pupils from lower socio-economic backgrounds had lower VRT scores. Pakistani pupils in this sample tended to come from lower socio-economic strata. VRT had a very strong effect on both KS3 and GCSE. Specifically with respect to sex and ethnic group differences, Pakistani pupils and girls underperformed academically relative to White pupils and boys at 14 years (KS3).

2.2. Model 2

A question not addressed in the foregoing analysis is whether there are SES, sex, or ethnic group differences in the six indicators over and above any differences in the two latent variables. In other words, it may well be the case that there are group differences in specific indicator variables, after we have taken into account whatever group differences may exist on the factor. Graphically, this situation is represented by parallel regression lines of unequal height. This question can be handled by inspecting the modification indices in model 1 above and allowing paths from the background variables directly into the indicators. Three such paths were freed up in this model. The first two concerned measurement intercepts for KS3 English, which were noninvariant across Pakistani and White UK pupils as well as across boys and girls (Fig. 1). The third path involved one measurement intercept for GCSE English, which was noninvariant across boys and girls.

The release of the relevant parameters led to a highly significant drop in the chi square value ($\Delta\chi^2_{(3)}=96.50$, $p<.01$). The completely standardized values for these paths are given in Fig. 1. Their substantive interpretations are that girls and Pakistani pupils perform better on the KS3 English assessment than would be expected on the basis of their standing on the KS3 factor. In addition, girls also perform better than expected on GCSE English.

3. Discussion

This study was designed to test two sets of hypotheses. First, it was hypothesised that different ethnic and gender groups would make different rates of progress through the secondary school years. On the basis of previous findings from cross sectional studies it was predicted that boys would progress less well than girls, low SES pupils would progress less well than high SES pupils, and Black and Pakistani pupils would progress less well than White UK pupils. The results of the MIMIC models show that Pakistani pupils and girls make poorer academic progress between the start of secondary school and assessment at KS3, relative to White pupils and boys, respectively. Secondly, it was hypothesised that the predictive validity of the verbal reasoning test scores would be similar across the different gender and ethnic groups. This was found to be the case in relation to the prediction of GCSE examination performance at age 16 years. However, some overestimation of female and Pakistani pupils' KS3 scores at 14 years was apparent.

With regard to the observed gender differences, it should be noted that the VRT is constructed so that equivalent mean scores are produced for boys and girls, which means that even larger differences may be obtained if KS2 scores, on which girls outperform boys (Office for Standards in Education, 1996) were used as the measure of performance at the end of primary school. The ethnic group differences suggest that omitting to control for SES or VRT, will exacerbate the academic performance discrepancy between Pakistani and White pupils, given that the former are disproportionately represented in the lower socio-economic strata and have significantly lower VRT scores.

In addition to the effects on the KS3 and GCSE latent variables, the MIMIC approach allowed for an investigation of possible direct effects from the background variables onto the six indicators. Three such significant effects were observed in this model. The direct effect from the Pakistani dummy variable on KS3 English suggests that Pakistani pupils perform better in the KS3 English exam than would be expected based on their overall performance in the KS3 assessment. It is possible that the subject of English is emphasized over and above the broader academic development of pupils from a Pakistani background, given that many of them are raised in homes were English is not the first language. The findings, however, show that this advantage does not carry over to the GCSE level, where Pakistani performance in English is consistent with their overall academic performance. The other two direct paths showed that girls performed better than expected both in KS3 English and in GCSE English. This result is fully in line with findings showing that verbal expression matures faster in girls than in boys and that scholastic achievement assessments involving writing, grammar, and spelling consistently tend to favour females (Jensen, 1998).

From a different viewpoint, the three direct effects discussed above indicate that KS3 English and GCSE English will be biased indicators of overall KS3 and GCSE performance. More specifically, these two indicators will overestimate overall KS3 and GCSE performance for girls. Likewise, KS3 English will over estimate overall KS3 performance for Pakistani pupils. Note that this type of bias is a function of the extent to which an indicator of

a latent variable has the same meaning across different groups of people (measurement invariance) and should be clearly distinguished from the concept of predictive bias in VRT scores.

The VRT is primarily a measure of crystallized cognitive ability, which concerns knowledge accumulated over a period of time and is heavily dependent on one's learning experiences (Cattell, 1971; Kline, 1991). The results show that there are reliable ethnic group differences in VRT scores, with Pakistani pupils scoring lower than their White UK counterparts. Similar findings have been observed in previous studies from the 1980s (West, Mascie-Taylor, & Mackintosh, 1992).

The importance of ethnic group differences in VRT scores lies in the fact that the test is a very strong predictor of subsequent academic performance at 14 and, later on, at 16 years. This is especially salient in Fig. 1, where we chose to let the residual variances of the latent variables correlate, instead of modelling a direct path from KS3 into GCSE, thereby bringing the strong effects of VRT to the foreground (for the latter analysis, see Petrides, Chamorro-Premuzic, Frederickson, & Furnham, 2005). It is worth noting that the size of these effects accords well with the empirical status of IQ as the strongest predictor of educational achievement (Jensen, 1998; Mackintosh, 1998; Sternberg, Grigorenko & Bundy, 2001). From a substantive perspective, it is interesting to note that, even after controlling for the very strong VRT effects, Pakistani pupils and girls underperformed at KS3, which suggests that factors other than crystallized cognitive ability contribute to the lower academic achievement of Pakistani pupils and girls at 14 years. These findings present a different picture of gender differences in academic performance to the widely reported data on proportions of boys and girls achieving five or more A to C grades in GCSE. They reinforce the importance of analysis by age and subject area and the need to take account of potentially confounding factors in interpreting differences in achievement.

Socio-economic status did not have significant effects on academic performance, although it was negatively correlated with VRT scores and the Pakistani dummy variable. It is possible that broader and more sensitive measure of SES would result in a comparatively small positive effect on achievement, but it is very unlikely that this effect would counterbalance the strong impact of cognitive ability or the ethnic group differences observed. Overall, these results are in accordance with North American data, showing that SES is much less important a determinant of educational attainment than cognitive ability (Herrnstein & Murray, 1994).

A significant practical implication of these results concerns the use of VRT scores in the context of educational selection and streaming. More specifically, the second hypothesis investigated whether the predictive validity of the test varies across sexes or across the three ethnic groups. A correct interpretation of the findings is that if VRT scores are used to predict academic performance at KS3 and GCSE, without reference to ethnic group membership or sex, then predictions concerning certain groups of pupils will be biased. The nature of the predictive bias is such that the achievement of girls and Pakistani pupils will be *over*estimated at KS3 level. In both cases, overestimation will be constant over the entire VRT score range. That is to say, girls and Pakistani pupils will consistently underperform relative to what would be expected of them on the basis of their VRT scores. No such bias seems to exist at the GCSE level.

The predictive bias of overall KS3 performance is mediated to the indicators of the latent variables. The higher the loading of an indicator on the factor, the greater the magnitude of the bias on that particular indicator will be. In most cases, however, the extent of the misestimation will be relatively small, in line with the size of the observed effects. This is all the more likely in the case of KS3 English because the direct effects from gender and the Pakistani dummy variable are in the opposite direction to the effects of these two variables on the KS3 factor. Thus, we would expect a common VRT regression line to underestimate the performance of boys on English (and overestimate the performance of girls), however, part of this misestimation would be counterbalanced by the direct (incremental) effect of gender on KS3 English, which would lead to an underestimation of the performance of girls (and overestimation of the performance of boys).

The foregoing discussion highlights some of the ways in which the data analytic techniques employed in this paper are superior to conventional analyses, such as multiple regressions. These include the great level of detail such techniques can provide and the specific advantages of the MIMIC approach in relation to modelling invariance across many subpopulations and assessing bias in observed indicators when sample sizes are small (Muthén, 1989). Other significant advantages of the general structural equation approach is that it can simultaneously estimate complicated models that would otherwise need a series of separate, yet partly overlapping analyses, it can allow for correlated residuals and it can take into account the measurement errors in the observed indicators of a latent variable. The last advantage is especially significant in cases where the interest is in assessing group differences in latent variables, because conventional analyses such as ANOVA, may detect spurious differences arising from error or specific variance in the indicators (Petrides, Jackson, Furnham, & Levine, 2003).

Certain limitations of this study must be acknowledged. The results speak only to possible biases with respect to academic performance as was operationalized in the present study. Different patterns of bias or indeed total absence thereof may be observed if different criteria are examined. Theoretically, it is possible that the predictive bias is entirely the result of measurement error in VRT scores (Jensen, 1980). This possibility could not be investigated here, as we did not have access to VRT item scores. It is also possible, however, that the bias lies on the side of the criteria rather than on that of the test. In addition, a caveat applies in relation to the small sample sizes of the ethnic groups (particularly Black participants). Last, it must be reiterated that the magnitude of the predictive bias in the VRT is small and unlikely to be of serious practical consequence.

In this study it was not possible to examine the academic performance of pupils from other ethnic backgrounds (e.g., Bangladeshi, Indian, and new immigrant groups) due to the small number of participants from these groups. Neither was it possible to differentiate precisely between the various Black groups that were combined for the purposes of this study (e.g., Black Africans, Black Caribbean, etc.). A larger sample would allow the models to

be expanded to incorporate additional background variables, such as type of school attended, which may have significant effects of their own or interact with other variables examined herein.

We would also acknowledge the debate about the use of broad measures of ethnicity as a predictor of achievement (Caldas, 1992; Penny & Bond, 1991). On the one hand, it is argued that race is not itself a proxy of any process and does not advance the understanding of educational achievement. On the other hand, it is argued that it is important to remove statistically the effect of race in order to identify variables that can inform the development of intervention programmes. On this basis, the examination of differences in attainment between different ethnic groups in different educational contexts would seem an important first step, leading to further investigation of factors influencing those differences that can be targeted for intervention.

These findings raise the possibility that the academic trajectories of many pupils may currently be determined before they take their KS3 exams at 14 years and long before they apply for university entry. Consequently, in addition to researching other process variables of interest, it seems important that future research covers the period preceding entry to secondary schooling. Concerns that increasing use of selection, streaming, and setting will affect the proportion of pupils from different ethnic groups being educated together appear to have validity. However, the results of this study indicate that the key equity issue concerns the relative educational outcomes obtained by different ethnic groups, not the relative validity of the measures used to predict those outcomes. This suggests the need for a research and policy focus on effective action to raise the achievement of lower performing groups, in particular during the years of primary schooling.

References

Caldas, S. (1992). Using multiple regression to predict achievement: A reply to Klingele and Warrick, Penny and Bond. *Journal of Educational Research*, *85*(6), 325−326.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Cole, N., & Moss, P. (1989). Bias in test use. In R. Linn (Ed.), *Educational measurement*, (3rd ed) New York: AERA/NCME, Macmillan.

Crook, D., Power, S., & Whitty, G. (1999). *The grammar school question: A review of research on comprehensive and selective school education.* London: Institute of Education Perspectives on Educational Policy Paper.

Demie, F. (2001). Ethnic and gender differences in academic achievement and implications for school improvement strategies. *Educational Research*, *43*, 91−106.

Department for Education and Skills (DfES) (2003). *Youth cohort study: The activities and experiences of 16 year olds: England and Wales 2002.* London: The Stationary Office.

Edwards, T., & Tomlinson, S. (2002). *Selection isn't working.* London: Catalyst Forum.

Gillborn, D., & Mirza, H. S. (2000). *Educational inequality: Mapping race, class and gender.* London: Ofsted.

Gillborn, D., & Youndell, D. (2000). *Rationing education: Policy, practice, reform amd equity.* Buckingham: Open University Press.

Haque, Z., & Bell, J. F. (2001). Evaluating the performance of minority ethnic pupils in secondary schools. *Oxford Review of Education*, *27*, 357−368.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hillard, A. G. (1994). What good is this thing called intelligence and why bother to measure it? *Journal of Black Psychology*, *20*, 430−444.

Howe, M. J. A. (1997). *IQ in question.* London: Sage.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1−55.

Ireson, J., & Hallam, S. (2001). *Ability grouping in education.* London: Paul Chapman Publishing.

Jensen, A. R. (1980). *Bias in mental testing.* London: Methuen.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*, 631−639.

Kline, P. (1991). *Intelligence: The psychometric view.* London: Routledge.

Laosa, L. M. (1996). Intelligence testing and social policy. *Journal of Applied Developmental Psychology*, *17*, 155−173.

Mackintosh, N. J. (1986). The biology of intelligence? *British Journal of Psychology*, *77*, 1−18.

Mackintosh, N. J. (1998). *IQ and human intelligence.* Oxford: Oxford University Press.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters.* London: Open Books.

Muthén, B. O. (1989). Latent variable modelling in heterogeneous populations. *Psychometrika*, *54*, 557−585.

Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide (v. 2).* Los Angeles, CA: Muthén & Muthén.

Nuttall, D., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, *13*, 769−776.

Office for Standards in Education (Ofsted) (1996). *The gender divide: Performance differences between boys and girls at school.* London: HMSO.

Penny, J., & Bond, L. (1991). Use of ethnicity as a predictor of achievement: A reply to Klingele and Warrick. *Journal of Educational Research*, *84*, 133−134.

Petrides, K. V., Chamorro-Premuzic, Frederickson, N., & Furnham, A. (2005). Explaining individual differences in scholastic behavior and achievement. *British Journal of Educational Psychology*, *75*, 239−255.

Petrides, K. V., Jackson, C. J., Furnham, A., & Levine, S. Z. (2003). Exploring issues of personality measurement and structure through the development of a revised short form of the Eysenck Personality Profiler. *Journal of Personality Assessment*, *81*, 271−280.

Petrides, K. V., Frederickson, N., & Furnham, A. (2004). The role of trait emotional intelligence in academic performance and deviant behavior at school. *Personality and Individual Differences*, *36*, 277−293.

Portes, P. R. (1999). Social and psychological factors in the academic achievement of children of immigrants: A cultural history puzzle. *American Educational Research Journal*, *36*, 489−507.

Rampton, A. (1981). *West Indian children in our schools.* London: HMSO.

Reschly, D. J., & Ward, S. M. (1991). Use of adaptive measures and over representation of black students in programmes for students with mild mental retardation. *American Journal of Mental Retardation*, *96*, 257−268.

Richardson, R., & Wood, A. (1999). *Inclusive schools, inclusive society: Race and identity on the agenda.* Stoke-on-Trent: Trentham Books.

Sammons, P. (1995). Gender, ethnicity and socio-economic differences in academic progress: A longitudinal analysis of student achievement over nine years. *British Educational Research Journal*, *21*, 465−485.

Social Exclusion Unit (1999). *Bridging the gap: New opportunities for 16–18 year olds not in education, employment or training.* London: The Stationary Office.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill Palmer Quarterly*, *47*, 1−41.

Swann, M. (1985). *Education for all: Report of the committee of inquiry into the education of children from ethnic minority groups.* London: HMSO.

Vernon, P. E. (1957). *Secondary school selection.* London: Methuen.

West, A. M., Mascie-Taylor, C. G. N., & Mackintosh, N. J. (1992). Cognitive and educational attainment in different ethnic groups. *Journal of Biosocial Science*, *24*, 539−554.