Institute of Education

# Does longitudinal change in adolescent educational expectations for university study vary by ethnic group?

**Michael Tzanakis**

CLS Working Paper 2015/04

June 2015

**Centre for Longitudinal Studies**
Following lives from birth and through the adult years
www.cls.ioe.ac.uk

CLS is an ESRC Resource Centre based at the UCL Institute of Education

# Does longitudinal change in adolescent educational expectations for university study vary by ethnic group?[1]

**Author: Michael Tzanakis, Department of Quantitative Social Science, Institute of Education, University of London**

**Abstract**:

Previous UK research suggests that adolescent expectations for university study are likely to be moderated by ethnicity. However, this conclusion needs to be qualified. We still need to assess the extent prior expectations affect later ones; whether they do so directly or indirectly; or whether differences in these effects across ethnic groups are statistically significant. To answer these questions, we need to explicitly address the measurement assumptions that past research has largely taken for granted. This quantitative study explores change in pupils' expectations between ages 14-16 systematically across white, Indian, Pakistani, Bangladeshi and Black Caribbean adolescents under a psychometric framework using cohort panel data from waves 1-3 of the Longitudinal Study of Young People in England (LSYPE). An autoregressive longitudinal latent variable structural equation mediation model enables estimation of direct and indirect effects and explicit tests of the assumptions of invariance, stationarity and equilibrium and measurement of group differences in longitudinal change in latent means and intercepts. Results show that pupils' educational expectations change dramatically differently across ages 14-16 in the five ethnic groups. Expectations at age 16 are affected directly and indirectly by prior expectations but ethnic groups differ systematically in these effects. Cross-group differences in latent means and intercepts suggest that in contrast to their white peers who exhibit greater stability in relatively lower expectations, minority pupils exhibit much higher expectations that are less stable because they increase over time. There is a general fall in expectations at age 15 but contrary to their white peers, all minority pupils recover at age 16.

**Keywords**: adolescent educational expectations, ethnicity, longitudinal mediation modelling, latent variable structural equation modelling

## 1. Introduction

Interest in adolescent educational expectations has had a long history in sociology (Krauss, 1964; Picou and Carter, 1976; Portes, McLeod and Parker, 1978; Sewell and Shah, 1968; Woelfel and Haller, 1971) and psychology (Gottfredson, 1981; Hofferth, 1980; Marjoribanks, 1999; Quaglia and Cobb, 1996). There has been an impressive array of both qualitative and quantitative work that has established the association of adolescent educational expectations with a variety of outcomes. It has been long established for example that higher educational

[1] The author would like to thank Professors Richard Wiggins, Alice Sullivan and John Micklewright, Department of Quantitative Social Science, Institute of Education for their valuable input in this study. An earlier version of this work was presented at the *16th Annual International Conference on Education*, 19-22 May 2014, Athens Institute for Education and Research, Athens Greece. The author expresses his gratitude to the Department of Quantitative Social Science, Institute of Education who funded his participation in the Conference.

expectations are associated with pupils' higher educational achievement as well as with better academic self-concept.

Yet, quantitative studies did not systematically study longitudinal change in academic expectations until recently when large-scale longitudinal panel data became available. Such studies have pointed to important shifts in educational expectations over the years and have therefore offered useful insights with regard to longitudinal change in expectations.

Important questions however remain. It is for example questionable that all young people change their expectations in the same way. It was indeed found that the trajectories of young people's educational expectations for university study changed differentially after age 14, depending on whether at age 14 these young people had reported very high, average or low likelihoods to apply to university or be successful in their application if they applied. Expressed in the language of mediation-moderation modelling (to be explained below), this finding suggests that level of expectations at age 14 moderated subsequent longitudinal change in expectations.

However, it is well known that both parental social class, particularly parental education and income as well as parental ethnicity are associated with differential rates of adolescent educational expectations. Recent analyses based on national samples have shown that higher parental social class is generally associated with higher adolescent educational expectations. This association is typically accounted for by the increased resources one would expect to find in higher parental social class, in particular, better capabilities resulting from higher income, social and cultural capital.

Only a small percentage of variation in adolescent educational expectations is typically accounted for by parental social class or its correlates, however. Parental ethnicity, on the other hand, upsets the expected association of parental social class and adolescent expectations. Adolescents from disadvantaged minority ethnic groups in the UK manifest significantly higher academic expectations as compared to their peers from relatively more advantaged families. This finding runs contrary to what one would expect based only on parental position. Although earlier research has brought attention to this paradox, ethnicity as a potential moderator of longitudinal change in adolescent expectations has not been systematically investigated. Further, the general statistical modelling approach (typically, OLS or logistic regression) adopted in past research has been limited in systematically estimating the significance of observed differences in the longitudinal change in adolescent expectations across UK ethnic groups. Further, differences in the longitudinal direct and indirect effects across several repeated measures of adolescent educational expectations remain unknown. The present study adopts a psychometric approach and studies longitudinal change of adolescent educational expectations of UK white, Indian, Pakistani, Bangladeshi and Black Caribbean adolescents under a multi-group analysis of a mediation-moderation latent variable longitudinal model. This approach enables precise measurement of *moderated mediation*. This involves estimation of the direct and indirect longitudinal effects in expectations between ages 14 and 16 as mediated via those at age 15. Systematic estimation of the significance of cross-group differences in the potential moderation of these effects by parental ethnicity is also undertaken.

Section 2 presents a brief literature review to showcase the paradox of high expectations for university study of adolescent pupils from relatively disadvantaged family backgrounds. The research questions are described in section 3 while the data source and treatment are described in section 4. Section 5 presents the methodological approach of estimating a longitudinal mediation / moderation model under a psychometric framework. Section 6 presents the results of the analysis and discusses them in more detail in Section 7 which also concludes.

## 2. Literature Review

### 2.1 Accounting for the paradox of adolescent educational aspirations, expectations and ethnicity.

Educational *aspirations* and *expectations* are related but theoretically distinct concepts (Quaglia and Cobb, 1996). In sociology both terms have been used (Alexander and Eckland, 1975; Kerckhoff, 1977; Sewell, Haller and Portes, 1969; Sewell, Hauser and Wolf, 1980). In psychology the term *aspirations* is preferred (Ritchie, Flouri and Buchanan, 2005). In theory, adolescent *aspirations* are argued to describe unrealistic or idealistic perceptions of future education or occupation (Goyette, 2008; Portes, McLeod and Parker, 1978; Woelfel and Haller, 1971). These may be based on fantasy rather than reality (Furlong and Biggard, 1999) but are considered normal during a natural developmental stage (Gottfredson, 1996). Occupational aspirations in particular are heavily gendered, and reflect a generally limited perception of surrounding market reality (Furlong, 1986; Kelly, 1989). *Expectations*, on the other hand, are realistic assessments or predictions of future attainments, job category (Woelfel and Haller, 1971) and available resources (Thompson, Alexander and Entwisle, 1988). This analysis is concerned with young people's reported *probability estimates* for applying to university after year 11 (age 16) and for being accepted if they applied. Pupils' responses are therefore more likely to reflect reality-based expectations rather than idealistic aspirations about post-16 university study. Accordingly, the term *expectations* is used in this study. Minority young people's expectations are likely to be shaped by a number of structural and cultural factors that may explain why their expectations are higher than those of their white peers. I review the evidence on such influences below.

It is well established that young people from poor backgrounds tend to have lower educational expectations and performances than their better-off peers (Hofferth, 1980; Leibnowitz, 1974; Murmane, Maynard and Otis, 1981; Rosen and Aneshensel, 1978; Sewell, Hauser and Wolf, 1980). More recent evidence has generally confirmed this association (Breen and Goldthorpe, 1997; Buchmann and Dalton, 2002; Jerrim, 2011; Sacker, Schoon and Bartley, 2002). UK ethnic minority groups vary in their socio-economic status, but most are relatively disadvantaged compared to whites (the main exception to this is Indian pupils). So, one might expect minority ethnic children to have lower educational expectations than whites, and to make less progress in school. Generally confirming quantitative research evidence based on large-scale nationally representative samples in both the US (Kao and Tienda, 1998; Qian and Blair, 1999) and Australia (Marjoribanks, 2003a; 2003b), precisely the opposite is the case in the UK as well (Tzanakis, 2014). This creates the paradox of high expectations among less advantaged ethnic minority youth in the UK. This inconsistency is typically duly noted but remains unexplored as to its causes.

Dominant theoretical perspectives in sociology are limited in explaining the above paradox. Influential perspectives in the UK sociology of education that involve educational expectations explicitly are rational action theory and the relative risk aversion mechanism (Breen and Goldthorpe, 1997; Goldthorpe and Breen, 2000); social capital (Coleman, 1988) and social and cultural reproduction (Bourdieu and Passeron, 1977). However, none of the above perspectives involves ethnicity per se in the formation of adolescent educational expectations. Thus, these perspectives cannot explain the paradox of high expectations for university study of pupils from relatively disadvantaged backgrounds, as is the case with most South Asian minority ethnic groups in the UK. South Asian parents' decisions to invest in their children's university education are not 'rational' in the sense predicted by rational action theory. Higher expectations of ethnic minority adolescents do not seem to follow class-induced cultural scripts, either (Goldthorpe, 2007). Another mechanism that promotes and maintains higher expectations in lower-income Indian, Pakistani and Bangladeshi families seems to operate less in white and Black Caribbean families. Addressing this point, Modood (2003; 2004) suggests that minority families possess orientations that are not defined by their relatively disadvantaged class position. Young members of minority groups may have deficits in class-defined cultural capital but they succeed academically by possessing peculiar variants of social and cultural capital (Basit, 2012; Siraj-Blatchford, 2010), not predicted by the above theories. These 'ethnic' capitals are argued to generate and maintain high academic expectations early on and promote higher academic performances.

The literature offers some explanation for this mechanism. We know that positive attitudes to school for example, are strongly associated with higher expectations to remain in education. These attitudes are in turn significantly associated with positive home learning environments, family socioeconomic status (SES) and individual characteristics (Marjoribanks, 2003a). Minority pupils' higher academic self-concept; peer support; and a higher commitment to schooling indicated by amount of homework and positive attitudes to school and teachers accounted for most of the variance in aspirations across UK minority ethnic groups (Strand and Winston, 2008; Stryker, 2007). Studies based on large-scale, nationally representative samples as the LSYPE (Strand, 2007; 2008) reported that the gaps in educational expectations between white British and the Indian, Pakistani, Bangladeshi and Black Caribbean pupils were explained mostly by home- and school-related factors rather than social class per se, offering support to previous studies such as Phillips (1998) and Sylva *et al.*,(2004). Recent research on expectations in the UK using the LSYPE panel data suggests that teenage expectations to apply to university or to be admitted if they apply are highly associated with pupils' early academic performance at age 11 (Anders and Micklewright, 2013); age 14 (Fumagalli, 2012) and age 15 (Croll and Attwood, 2013).

Academic performance as indicated by earned grades has been routinely used as a proxy for ability, in the absence of direct IQ measurements (Bond and Saunders, 1999). However, the direction of causality between academic performance and educational expectations is unclear. Even if we consider grades as acceptable proxies for ability, it is questionable that the better academic performances of UK ethnic minority groups as compared to those of their white peers are due to differentials in innate ability. Research regarding the academic performance of UK ethnic minorities has shown that in terms of overall progress between ages 7 and 11, Black Caribbean, Black Other and Pakistani pupils are reported to have progressed less and Bangladeshi and Chinese pupils more than their white peers (DfES, 2006; Melhuish *et al.*, 2006; Modood, 2005; Strand, 1999). However, the trend is reversed between ages 11 and 16 (Wilson, Burgess and Briggs, 2005). All minority ethnic pupils are reported to have made much more progress than their white peers in post-secondary education (Modood, 2003; 2004; Stevens, 2007). Given that reversal, it is hard to accept that the observed ethnic differences in academic performance involve group differentials in ability. The same argument can be made between the potential association of ability, grades and educational expectations. The paradox of young people's high expectations from low SES ethnic minority families and thus, longitudinal changes in such expectations, are hard to explain solely on the basis of ability, grades or parental social class although there is evidence that the three factors are related (Feinstein, 2003; Feinstein, Duckworth and Sabates, 2004). Recent research by Ross and Lloyd (2013) found that the increases in the cost of university education in the UK were a strong deterrent for at least one third of pupils who at age 14 thought it very or fairly likely to apply to university. Thus, living in a family of £26,000 or lower increased a young person's odds of being 'concerned' by 40% (Ross and Lloyd, 2013, p.37). Yet, pupils from typically disadvantaged UK minorities like Pakistani, Bangladeshi including Indian young people from lower-income families had the *lowest* odds of being concerned or deciding against university education at age 20/21 (*ibid*, p. 63). This suggests that the teenage expectations of disadvantaged UK ethnic minorities are atypical and should be studied separately rather than as part of a general sample of young people.

### 2.2 Longitudinal change in adolescent expectations

Recent quantitative research in the UK and the USA has focused on the longitudinal change in adolescent educational expectations based on large-scale, nationally-representative datasets. Based on the LSYPE waves 1 to 7, Croll and Attwood (2013) found that early adolescent educational expectations at age 14 predicted actual applications at age 20/21. However, although parental SES matters, it is a much weaker influence on pupils' adolescent expectations about actual applications for university study than pupils' good prior academic performance even as early as age 11. Anders and Micklewright (2013) who used the same LSYPE data found that expectations started lower and fell faster for those teenagers from disadvantaged backgrounds, or those who left full-time education (FTE) at age 16. By contrast, high expectations persisted over time for all those who remained in FTE and did A-levels. However, while 66% of teenagers with high SES parents and high KS2 performances actually applied at age 20/21, 50% of teenagers with low SES but similarly high KS2 performances also did the same. Even though this appears like a significant difference, the point remains that expectations appear to be associated more with actual performance and much less with SES. Working with the same data, Fumagalli (2012) showed that teenage expectations evolved the most during the transition between age 14 to 15 (years 9 to 10) as a consequence of new information received after the results of pupils' KS3 performance tests. The positive effect of KS3 performance at age 14 on expectations to be admitted to university recorded at age 15, was stronger in the case of pupils from high parental SES with high-educated parents and weaker in the case of their peers from disadvantaged backgrounds. However, even controlling for KS3 performance, teenagers from disadvantaged homes held similarly high expectations of being admitted to university. It is therefore plausible to expect that for Indian, Pakistani and Bangladeshi adolescents who have typically better academic performances than their white and Black Caribbean peers at Key Stage 2-3 and GCSE tests (Modood, 2003, 2004), parental SES will be a rather weak longitudinal influence on their expectations for university study. When this hypothesis was subjected to rigorous tests using the same data, it was found that parental social position at pupils' age 14 had a positive but generally negligible longitudinal effect on pupils' expectations at age 16 (Tzanakis, 2014).

Jacob and Wilder (2010) also studied trends in educational expectations of US adolescents between the mid-1970s and the early 2000s using the High School and Beyond, the 1988 National Educational Longitudinal Survey and the Monitoring the Future longitudinal datasets. They found that young people's educational expectations rose for all students during the analysis period with young women showing the greatest increases, confirming earlier evidence (Kao and Thompson, 2003; Kao and Tienda, 1998; Qian and Blair, 1999). Just as in the case of UK, family SES in the US tended to become less predictive of student expectations over time while in the case of young women, it was 'virtually uninformative' (Jacob and Wilder, 2010, p. 20). Academic performance remained a relatively strong predictor of future expectations but expectations also tended to become less predictive of future attainment.

**3. Research Questions**

The research questions in this study were the following:

a. Do expectations remain similar constructs over ages 14-16? (testing the assumption of longitudinal measurement invariance between ages 14 to 16)
b. Do adolescent academic expectations change significantly between ages 14 to 16? (testing the assumption of stability (i.e., longitudinal structural invariance) in expectations over time)
c. Are expectations at ages 15 and 16 affected similarly by their prior expectations at ages 14 and 15? (testing the assumptions of stationarity and equilibrium in expectations over time)
d. Are expectations at age 16 affected directly and indirectly by prior expectations at ages 15 and 14? (testing the hypothesis of mediation of the effect of expectations at age 14 on those at age 16 via expectations at age 15)
e. Are expectations similar constructs at ages 14, 15 and 16 across UK minority ethnic groups? (testing the assumption of cross-group measurement invariance of educational expectations across the UK ethnic groups in the study)
f. Is stability in expectations between ages 14 to 16 significantly different across UK minority ethnic groups in the study? (testing the hypothesis of moderation of longitudinal change by maternal ethnic group membership)
g. Are the above direct and indirect effects significantly different across UK minority ethnic groups? (testing the hypothesis of moderation of the direct and indirect effects by maternal ethnic group membership, i.e., the assumption that the potential mediation of the effect of expectations at age 14 on those at age 16 via expectations at age 15 is moderated by maternal ethnic group membership)

**4. Data source and treatment**

The data source for this study is the Longitudinal Study of Young People in England (LSYPE). The LSYPE is a multipurpose, ongoing large-scale panel study of young people in England commissioned by the Department of Children, Schools and Families (DCSF). The study followed a cohort of **15,770** young people who were in year 9 between ages 13-14 in maintained and independent schools and pupil referral units in England on February 2004. It has followed a complex, two-stage random probability, proportional to size (PPS), sampling design with disproportionate stratification. Schools were the primary sampling units. Deprived schools and pupils from the major UK minorities (Indian, Pakistani, Bangladeshi, Black Caribbean) were oversampled by a factor 1.5. Table 1 shows the distribution of young people by gender and ethnic group between waves 1-3 prior to treatment for the needs of this study.

**Table 5.2: Young Person's self-reported ethnic group (grouped) by gender, Waves 1-3.**

| | Wave1 | | | Wave2 | | | Wave3 | | |
|---|---|---|---|---|---|---|---|---|---|
| YP' ethnic group | Male | Female | Total | Male | Female | Total | Male | Female | Total |
| White | 5343 | 4992 | 10335 | 4488 | 4427 | 8915 | 4149 | 4059 | 8208 |
| % | 51.7 | 48.3 | 100.0 | 50.3 | 49.7 | 100.0 | 50.5 | 49.5 | 100.0 |
| Mixed | 383 | 416 | 799 | 377 | 324 | 701 | 333 | 326 | 659 |
| % | 47.9 | 52.1 | 100.0 | 53.8 | 46.2 | 100.0 | 50.5 | 49.5 | 100.0 |
| Indian | 529 | 484 | 1013 | 433 | 417 | 850 | 401 | 388 | 789 |
| % | 52.2 | 47.8 | 100.0 | 50.9 | 49.1 | 100.0 | 50.8 | 49.2 | 100.0 |
| Pakistani | 468 | 472 | 940 | 400 | 402 | 802 | 369 | 381 | 750 |
| % | 49.8 | 50.2 | 100.0 | 49.9 | 50.1 | 100.0 | 49.2 | 50.8 | 100.0 |
| Bangladeshi | 322 | 401 | 723 | 322 | 301 | 623 | 292 | 272 | 564 |
| % | 44.5 | 55.5 | 100.0 | 51.7 | 48.3 | 100.0 | 51.8 | 48.2 | 100.0 |
| Black Caribbean | 287 | 289 | 576 | 254 | 252 | 506 | 233 | 235 | 468 |
| % | 49.8 | 50.2 | 100.0 | 50.2 | 49.8 | 100.0 | 49.8 | 50.2 | 100.0 |
| Black African | 296 | 317 | 613 | 260 | 272 | 532 | 248 | 244 | 492 |
| % | 48.3 | 51.7 | 100.0 | 48.9 | 51.1 | 100.0 | 50.4 | 49.6 | 100.0 |
| Other | 219 | 199 | 418 | 195 | 172 | 367 | 175 | 170 | 345 |
| % | 52.4 | 47.6 | 100.0 | 53.1 | 46.9 | 100.0 | 50.7 | 49.3 | 100.0 |
| Total | 7847 | 7570 | 15417 | 6729 | 6567 | 13296 | 6200 | 6075 | 12275 |
| % | 50.9 | 49.1 | 100.0 | 50.6 | 49.4 | 100.0 | 50.5 | 49.5 | 100.0 |

Source: LSYPE Young Person Files (Longitudinal), Waves 1-3
Note: Data are unweighted to show unadjusted percentages and frequencies. Totals in the last row exclude missing cases.

There was a very good response rate between LSYPE waves 1-3 – see Table 2

**Table 2: Overall response by LSYPE Wave 1-3**

| Sample | Issued | Households reached (%)* | Fieldwork period | Full Interviews | Partial Interviews |
|---|---|---|---|---|---|
| Final | 21,447 | - | | - | - |
| Wave 1 | 15,770 | 15,770 (100.0) | 30/3 – 18/10/2004 | 13,914 | 1,856 |
| Wave 2 | 15,678 | 13,539 (86.0) | 18/4 – 18/09/2005 | 11,952 | 1,587 |
| Wave 3 | 13,525 | 12,439 (90.0) | 21/4 – 28/09/2006 | 12,148 | 291 |

Source: NatCen, 2009 *Percentages in parentheses refer to percent household reached based on achieved sample base from previous wave.
Note: Data are unweighted to show unadjusted frequencies. Full interviews: YP, MP and SP interviewed; Partial interviews: not all members of the household interviewed

The main interest of the present analysis is to assess the extent to which maternal ethnicity moderates longitudinal change in pupils' expectations for university study between ages 14-

16. It considers maternal ethnicity as a *distal* moderator of individual-level proximal processes (Bronfenbrenner, 2005). This information was contained in mother's ethnic group membership as recorded at waves 1 to 3 of the LSYPE. The analysis required therefore repeated measures contained in the LSYPE waves 1-3 for the white, Indian, Pakistani, Bangladeshi and Black Caribbean groups. The categories 'mixed' and 'other' in both mothers and young people were excluded as their ethnicity was indeterminate. Black African mothers and young people were also excluded because of particularly high levels of item missingness and wave 1-3 non-response. The derived file included **10,915** mothers as well as the young people clustered under the same household identified by the unique survey identification code. The modelling assumptions discussed below required that the repeated measures included in the analysis represented responses from the same persons only. In the LSYPE these are identified as 'main parents' (about 90% of whom were mothers) and 'young persons'. As a result, a subset of the derived file was created that contained only fully-productive mothers (i.e., mothers and young people who had consistently participated in all 3 LSYPE waves). This produced the master working file containing **10,633** mothers and the young people structure clustered under the same household survey identification code.This wave 1-3 longitudinal file was stratified by five pre-selected mothers' ethnic groups. Table 3 shows the initial distribution or these pre-selected or trimmed samples.

**Table 3 Sample sizes of mothers' groups**

| Groups | N |
|---|---|
| White | 7578 |
| Indian | 751 |
| Pakistani | 642 |
| Bangladeshi | 484 |
| Black Caribbean | 324 |

Source: LSYPE waves 1-3 longitudinal File

The present analysis used repeated measures of the same two LSYPE variables regarding adolescent expectations that were used by Anders and Micklewright (2013), Croll and Atwood (2013) and Fumagalli (2012). Young people at LSYPE waves 1-3 were first asked to give an estimate of the likelihood to apply to university: '*How likely do you think it is that you will ever apply to go to university to do a degree?*' which ranged from 'very likely', 'fairly likely; 'not very likely' and 'not at all likely'. Young people who selected the first three responses above were subsequently asked to give an estimate of the likelihood to be admitted to university if they applied: '*How likely do you think it is that if you do apply to go to university you will get in?*', which had a similar response range identical to the first question. Both questions included a 'don't know' response which was treated as a missing value in the present analysis. Those young people who had selected the 'not at all likely' option in the first question were treated as missing values in the second question along with those who selected the 'don't know' option. Tables 4 and 5 show the unadjusted frequencies prior to multiple imputation of missing values for the two variables.

**Table 4: Pupil-reported likelihood to apply to university by ethnic group, at LSYPE waves 1-3**

| | Wave 1 | | | | | Wave 2 | | | | | Wave 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level of likelihood | White | Indian | Pakis | Bang | BC | White | Indian | Pakis | Bang | BC | White | Indian | Pakis | Bang | BC |
| 1.00 Not at all likely | 12.0 | 2.2 | 3.2 | 4.1 | 4.1 | 15.4 | 3.0 | 3.6 | 5.1 | .4 | 19.9 | 3.3 | 6.2 | 4.6 | 1.5 |
| 2.00 | 18.6 | 4.6 | 8.6 | 11.4 | 11.4 | 20.6 | 5.7 | 9.1 | 9.8 | 6.3 | 18.5 | 3.7 | 6.4 | 9.2 | 5.5 |
| 3.00 | 36.7 | 34.5 | 40.2 | 41.9 | 41.9 | 32.9 | 29.7 | 36.5 | 38.9 | 68.6 | 25.9 | 23.7 | 36.2 | 36.7 | 60.4 |
| 4.00 Very likely | 32.7 | 58.8 | 48.1 | 42.6 | 42.6 | 31.1 | 61.6 | 50.7 | 46.1 | 24.7 | 35.7 | 69.3 | 51.2 | 49.6 | 32.7 |
| Total | 7578 | 751 | 642 | 484 | 324 | 7578 | 751 | 642 | 484 | 324 | 7578 | 751 | 642 | 484 | 324 |
| Missing | 4.9 | 3.5 | 7.3 | 9.3 | 9.3 | 4.9 | 2.3 | 6.1 | 7.6 | 16.4 | 4.5 | 2.3 | 4.5 | 5.4 | 15.1 |

Note: Pakis=Pakistani; Bang=Bangladeshi; BC=Black Caribbean

**Table 5: Pupil-reported likelihoods of being accepted to university if applied by ethnic group at LSYPE waves 1-3**

| | Wave 1 | | | | | Wave 2 | | | | | Wave 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level of likelihood | White | Indian | Pakis | Bang | BC | White | Indian | Pakis | Bang | BC | White | Indian | Pakis | Bang | BC |
| 1,00 Not at all likely | 1.7 | .6 | .9 | .5 | .4 | 2.1 | .9 | 1.1 | .8 | .4 | 1.9 | .1 | .2 | .5 | 1.5 |
| 2,00 | 15.2 | 4.6 | 5.7 | 9.2 | 9.2 | 16.7 | 3.8 | 7.7 | 8.0 | 6.3 | 13.1 | 3.4 | 4.5 | 7.8 | 5.5 |
| 3,00 | 64.2 | 60.7 | 58.7 | 59.2 | 61.0 | 60.3 | 57.9 | 59.0 | 56.5 | 68.6 | 57.9 | 53.6 | 56.1 | 58.3 | 60.4 |
| 4,00 Very likely | 19.0 | 34.0 | 34.6 | 31.1 | 29.4 | 20.9 | 37.4 | 32.2 | 34.8 | 24.7 | 27.1 | 42.9 | 39.2 | 33.5 | 32.7 |
| Total | 7578 | 751 | 642 | 484 | 324 | 7578 | 751 | 642 | 484 | 324 | 7578 | 751 | 642 | 484 | 324 |
| Missing | 21.8 | 10.8 | 15.4 | 19.0 | 16.0 | 25.0 | 9.9 | 14.5 | 17.4 | 16.4 | 28.1 | 9.3 | 17.3 | 14.9 | 15.1 |

Note: 1.00=not at all likely; 2:00=not likely; 3.00=likely; 4.00=very likely

The tables show that more than one third of Indian, Pakistani, Bangladeshi and slightly less of Black Caribbean young people expected to go to university and considered successful application for university study ‘very likely’ as compared to one fifth of their white counterparts. Although the proportions of white pupils who regarded a successful university application very likely increased in year 16, the proportions of all the other minority groups increased even more. So, the gaps in educational expectations remained.

However, the sample sizes of the white and the minority ethnic groups were widely discrepant. Sample size ratios between the white and the other ethnic groups ranged between a maximum of 23.4 and a minimum of 10.1. Aguinis (1995, p. 1148) demonstrated that if the moderator (maternal ethnicity in this case) is measured by several ethnic groups of unequal sample sizes, the power to detect moderation effects is severely reduced in multiple regression models. An analogous but more serious issue emerges when ethnicity groups of widely discrepant sample sizes are included in the same multigroup analysis involving latent variable SEM as is the case in this study (Little, 1997; Vandenberg and Lance, 2000) (see below). Detection of moderator effects in multigroup analysis involving SEM rests upon the correct execution and interpretation of invariance tests. However, when widely discrepant sample sizes are included in the same multigroup analysis, severe bias results in the multigroup chi-square, the primary index of overall model fit and chi-square difference tests (Brown, 2006, p. 279). Also, modification indices, standard errors, power to detect parameter estimates as significantly different from zero and error variances will be differentially impacted by the unbalanced group sizes (Kaplan and George, 1995). Therefore, Type I error rates will be inflated because the null hypothesis that groups have equivalent structures will be rejected more often (Chen, 2007; 2008) To minimize bias in cross-group invariance tests under the present sample structure, I reduced the initial sample of the white group to balance the group sample sizes. Unfortunately, there are no specific guidelines in the methodological literature as to which ratio (apart from the ideal 1:1) between the reference and comparison group sample sizes is optimal to minimize such bias (Byrne and Stewart, 2006). Therefore, I considered the 4/1 ratio used in the simulation studies by Chen (2007; 2008), as the lower bound of permissible sample size discrepancy.

Since the smallest sample size was 324, the sample size of the white mothers was reduced to 1000 cases by drawing a random sample out of the original 7,578 cases. This was a 13.2% random sample of the original representative sample. This reduction did not affect the magnitudes of measurement or structural estimates other than their standard errors[2].

All missing values in the white group (n=1000), Indian (n=751), Pakistani (n=642), Bangladeshi (n=484) and Black Caribbean (n=324) groups were imputed using the MCMC algorithm in SPSS version 20 (IBM, 2011). A separate analysis of missingness before the imputation procedure was carried for each ethnic group. The longitudinal weight provided by the LSYPE was incorporated into the imputation procedure. The data showed a Missing At Random (MAR) pattern. Following the imputation of missing values, weighted covariance matrices were produced for each ethnicity group, which were used as input for subsequent multigroup analyses. The SEM software used in the multigroup analysis was AMOS Graphics 20 (Arbuckle, 2011).

## 5. Empirical strategy and method of analysis

Anders and Micklewright (2013), Croll and Atwood (2013), Fumagalli (2012) and Strand and Winston (2008) filled an important gap in the UK research on educational expectations. However, none of these researchers offered a formal test of the significance of the observed differences in educational expectations over time nor of ethnicity as a potential moderator of these differences. Further, all of the above longitudinal studies utilised multiple or logistic regression designs which assume that the observed variables are measured without error (Jaccard and Wan, 1995) with all types of error clustered in a single term. These designs also assume that the observed repeated measures used to study expectations longitudinally remain identical over time, which is an untested assumption. For example, young people's identical responses (i.e., answering 'very likely') to applying to university or of being admitted if they apply across ages 14, 15 and 16 might represent true responses. However, these responses may also carry a degree of common method or trait variance that may be the result of a respondent's remembering his or her previous response and repeating it in a later response, desirability bias, interviewee cultural predispositions or the interviewer's similar method of eliciting the response to or translation of the question. This potential unobserved common variance, which in the above designs is clustered with all measurement error, must be extracted before true change over time in expectations is assessed. Further, Anders and Micklewright (2013), Croll and Atwood (2013), Fumagalli (2012) who used the same LSYPE data, opted either for the repeated measures of the young people's reported probability of

[2] Preliminary analytical checks that were carried out in advance of the modeling showed that differences in the size of factor loadings and indicator intercepts were trivial when the model was estimated on the complete sample of white mothers (n=7578) and then on the 1000 case-sample. These results were consistent with simulation studies which showed that model parameters were generally unaffected by smaller sample sizes (Finch, West and MacKinnon, 1997). Full details of these checks are available on request from the author.

applying to university or for the reported probability to be admitted if they apply, which was nested within the first, but not both (see below). Fumagalli (2012) argued that responses to the former question represented *aspirations*, while responses to the latter, more concrete *expectations* regarding university entry. Clearly however, this is an arbitrary interpretation, as both variables are highly correlated (Anders and Micklewright, 2013; Croll and Attwood, 2013). Consequently, they are likely to represent two related dimensions of the same underlying construct. Failing to include both in a longitudinal design may leave out an important dimension in a measure of young people's expectation to attend university. Finally, none of the above approaches to studying longitudinal change allows for studying indirect as well as direct longitudinal effects of prior expectations on later expectations. Nor, can they test whether the yearly time lags representing change between a previous and a next LSYPE wave are statistically equivalent, another assumption that remains typically untested. Finally, assumptions about longitudinal or cross-group equivalence (invariance) of the observed measures of adolescent expectations were also untested.

### 4.1 The longitudinal model of adolescent expectations in this study

The longitudinal model of the present analysis addresses all the above assumptions explicitly as I explain in section 4.2 below. The model studies change in adolescent expectations between ages 14 (year 9) to 16 (year 11) across separate groups of white, Indian, Pakistani, Bangladeshi and Black Caribbean young people. The five groups were stratified according to the ethnic group to which their mothers, rather than themselves, belonged because this study is interested in measuring the distal influence of maternal rather than young people's own ethnicity. Although a time window of only two years is relatively short, years 9-11 represent a crucial period in the lives of UK adolescents. Year 11 marks the end of compulsory education in England and the year where the decision to stay on in FTE or enter the job market materialises. It is also the period that adolescents and their parents are likely to enter evaluating the pupils' prior academic performance at KS2 examinations sat at age 11 (Anders and Micklewright, 2013) as well as information from KS3 (Fumagalli, 2012) examinations sat at age 14. Finally, based on evidence gleaned from GCSE examinations (Croll and Attwood, 2013) lasting between ages 15 and 16, adolescents are expected to crystallise their expectations regarding staying on in FTE and applying to university.

For each of the five groups in the analysis, I treat adolescent expectations as a latent unobserved dimension represented by a set of observed (or manifest) variables which are imperfect measures of their underlying construct. *Latent variables* are unobserved constructs which are measured only via a set of observed variables, called *indicators* (or *manifest variables*) for which there are collected data. In this analysis, these indicators are represented by the repeated measures of the two LSYPE questions whose unadjusted frequencies are shown in Tables 4 and 5. Both questions were therefore four-point ordinal Likert type questions and were treated as continuous following general practice. Simulation studies have shown that the assumption or continuity holds if ordinal-level variables have $\geq 5$ categories (Babakus, Ferguson and Jöreskog, 1987; Bentler and Chou, 1987; Byrne, 2010; Muthén and Kaplan, 1985). Both indicators were very close to this range. As already referred to above, both questions were highly and consistently intercorrelated across waves 1-3 (Anders and Micklewright, 2013). So, it is plausible that their intercorrelation is due to their dependence on the underlying latent construct of adolescent educational expectations for university study. If, in other words, the influence of the latent variable on the indicators was partialled out, the correlation among the indicators would be zero (Bollen, 1989a).

It is assumed that the unobserved construct exists separately from its observed measures without contradicting observed data (Raykov and Marcoulides, 2006). Latent variables (or factors) are considered superior to observed (manifest) measures because they minimize measurement error (Muthén and Asparouhov, 2014). Instead of assuming that each observed variable in a multiple regression is measured without error, this assumption is explicitly tested in a CFA model. The variance of each measured indicator is partitioned into the variance explained by the latent factor (also called *communality*) and the variance which is accounted for by measurement error or other unobserved influences (called ‘*unique’* variance, ‘*error variance’* or simply ‘*uniqueness’*). Partitioning the indicator variance in this way allows the researcher to test hypotheses about potential relationships among indicator uniquenesses in a CFA model with multiple latent constructs. Thus, common variance due to respondents’ systematic similarities in their responses due to like mindsets or common methods in eliciting responses to multiple-choice questions (common *method variance*) or shared traits (common *trait variance*) can be extracted in a CFA model.

A CFA model possesses two parts: a *measurement model* which includes the estimates of the indicator loadings, the indicator errors, and the covariances among the indicator errors. If there are no covariances hypothesized among the indicator errors, it is assumed that such covariances are constrained to zero. When these covariances are specified, they are unconstrained and freely estimated. The hypothesized relations among the error terms of a CFA model are called ‘error *theory*’ or ‘*error structure’* (Kline, 2005). The second part, called the *structural model*, includes the relations among the latent constructs themselves. Typically, in a CFA model these are factor covariances. When a CFA model is extended to become a latent variable structural equation model (LV-SEM), the structural model includes all hypothesized regression paths among the independent (called *exogenous*) and the dependent (called *endogenous*) variables. A LV-SEM is confirmatory in that it tests a postulated structure informed by causal hypotheses (Raykov and Marcoulides, 2006). The LV-SEM of the present study appears in Figure 1, below:

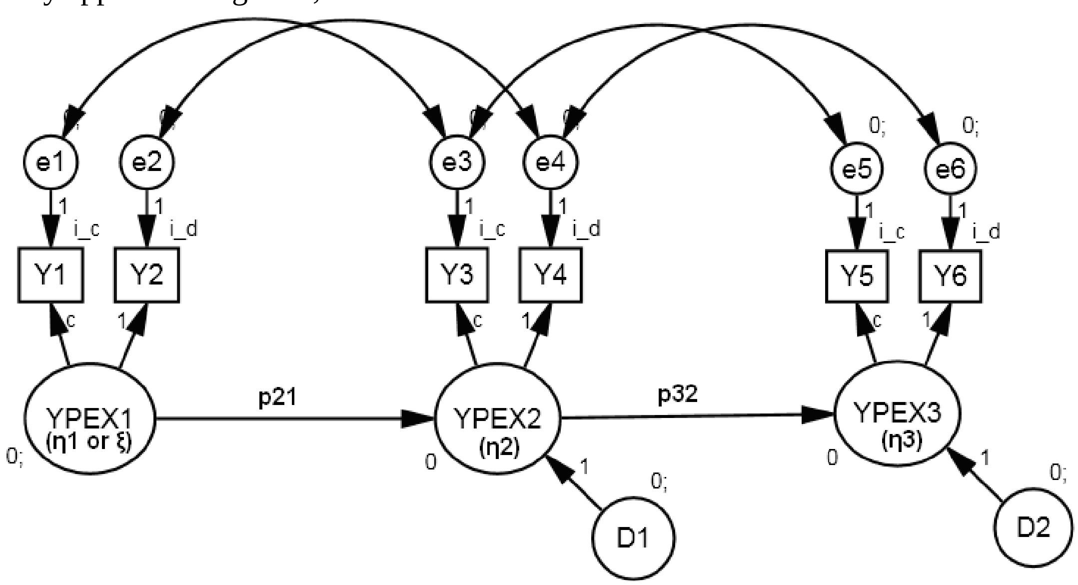


**Figure 4.7: The autoregressive model for adolescent educational expectations (YPEX) for ages 14 (YPEX1), 15 (YPEX2) and 16 (YPEX3) with identification, metric, scalar and longitudinal structural invariance constraints in place.**

**Legend:** Y1, Y3, Y5 (reported likelihood to apply to university at ages 14, 15 and 16); Y2, Y4, Y6 (reported likelihood to be admitted if apply at ages 14, 15 and 16, LSYPE waves 1-3).

In the above LV-SEM, each latent factor represents the same construct at various consecutive measurement points. For each occasion, there is a CFA model with its own variance-covariance matrix **S**. The assumption that these matrices, $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_n$ are invariant can be explicitly tested. Figure 1 illustrates a longitudinal autoregressive LV-SEM with three repeated measures (occasions) representing each construct and an autocorrelated error structure. These models are called 'autoregressive' latent variable SEM, because each occasion-specific latent construct is regressed onto its temporally prior occasion-specific latent construct (see, Bijleveld *et al.*, 1998, p. 4) to test specific causal hypotheses (Shrout and Bolger, 2002). In the present study, the causal hypotheses concern the relation between pupils' expectations at age 14 (YPEX1) and age 15 (YPEX2) and between those at age 15 and those of age 16 (YPEX3). These relations are also called *dependence* paths (Bijleveld *et al.*, 1998) because they describe the time dependence of an occasion on its temporally prior one. Thus, in its measurement part the model presented in Figure 1 estimates six factor loadings of observed variables Y1-Y6 regressed on their respective latent constructs, six unique error variances (e1-e6) and four error covariances ($cove_1e_3$, $cove_2e_4$, $cove_3e_5$, and $cove_4e_6$). In its structural part, the model estimates two structural paths ($p_{21}$ and $p_{32}$) and two disturbance terms (D1 and D2). To achieve identification, all factor means, factor intercepts and error means are fixed to zero, and the loadings of Y2, Y4 and Y6 as well as all the loadings of indicator errors and disturbance terms have been fixed to unity (1).

Figure 1 suggests a *congeneric* factor structure (Jöreskog, 1969; 1971), i.e., that each latent variable is defined only by its two indicators without any cross-loadings. X notation pertains when the construct is used as an exogenous factor while Y notation is used when the construct is an endogenous factor (see, Brown, 2006, p. 55-6). Using Y notation for the moment, the relations of the indicators to their construct are given by the equations 1-6 below, under the assumptions of normally-distributed, homoscedastic errors with zero means which are uncorrelated with the latent factors $\boldsymbol{\eta}_2$ , $\boldsymbol{\eta}_2$ , $\boldsymbol{\eta}_3$ (in this case represented by YPEX1-3); univariate normal and multivariate normal y indicators (Tabachnick and Fidell, 2001). The assumption of multivariate normality could be tolerated as the skew and kurtosis of both indicators at ages 14-16 was well within the limits typically suggested in the literature (see below). However, the assumption of uncorrelated residuals among themselves can be directly tested in the model of Figure 1 by testing the significance of the freely-estimated error covariances.

$$y_1 = \lambda_1\boldsymbol{\eta}_1 + \varepsilon_1 \qquad (1)$$
$$y_2 = \lambda_2\boldsymbol{\eta}_1 + \varepsilon_2 \qquad (2)$$
$$y_3 = \lambda_3\boldsymbol{\eta}_2 + \varepsilon_3 \qquad (3)$$
$$y_4 = \lambda_4\boldsymbol{\eta}_2 + \varepsilon_4 \qquad (4)$$
$$y_5 = \lambda_3\boldsymbol{\eta}_3 + \varepsilon_3 \qquad (5)$$
$$y_6 = \lambda_4\boldsymbol{\eta}_3 + \varepsilon_4 \qquad (6)$$

The intercept in equations 1-6 is suppressed because typically, indicator intercepts and means are not included in the analysis of covariance structures in CFA (Blunch, 2010; Brown, 2006). However, the present analysis will include means and covariance structures (MACS) (Bowers *et al.*, 2010; Byrne and Stewart, 2006; Hertzog and Schaie, 1988; Jöreskog, 1974; Little, 1997; Widaman and Reise, 1997) because a covariance matrix including a vector of indicator means is also supplied as input (called *augmented variance-covariance matrix*). Equations 1 – 6 can be summarized in matrix form:

$$y=\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (7)$$

which expands and generalises to

$$\boldsymbol{\Sigma}_y = \boldsymbol{\Lambda}_y\boldsymbol{\Phi}_{\eta\eta}\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_y \quad (8)$$

.

However, because the first occasion of the autoregressive LV-SEM of Figure 1 above is exogenous, it will be symbolised from now on by $\boldsymbol{\xi}$ (rather than $\boldsymbol{\eta_1}$**)** while the second and third occasions are endogenous and are symbolised by $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_3$ to which the disturbance terms *D1* and *D2* are attached. These terms capture all the variance in $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_3$ not explained by their prior occasions (which is $\boldsymbol{\xi}$ for $\boldsymbol{\eta_2}$ and $\boldsymbol{\eta_2}$ for $\boldsymbol{\eta_3}$). Thus, the following matrix equations express the relations of the $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ factors to their respective measurement models.

$$y=\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (9)$$

$$x=\boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \quad (10)$$

where y and x code the vectors *p* of the indicators of endogenous and exogenous latent constructs and $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ the *p*p* matrices of x and y loadings ($\lambda$ of $\boldsymbol{\xi}$ and $\lambda$ of $\boldsymbol{\eta 2}$ and $\boldsymbol{\eta 3}$ in Figure 1). If *m* codes the number of endogenous and *n* the number of exogenous factors, $\boldsymbol{\eta_2}$ and $\boldsymbol{\eta_3}$ represent an *m*m* and $\boldsymbol{\xi}$ an *n*n* matrix of the endogenous and exogenous latent factors, while $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ the *p*p* matrices of indicator uniquenesses of the $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ factors. The original specification of a structural equation model was described as 'Linear Structural Relationships' or LISREL and defined algebraically(Jöreskog, 1969; 1993; Jöreskog and Sörbom, 1984) as:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (11)$$

with $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ defined as above and $\boldsymbol{\zeta}$ representing unexplained variability (error). **B** is an *m*m* matrix describing the relations among the endogenous latent variables and the $\boldsymbol{\Gamma}$ an *m*n* matrix representing the relations between any endogenous (*m*) and exogenous (*n*) latent variables. In addition, an $\boldsymbol{\Phi}$ covariance (*n*n*) matrix specifies the covariances among any exogenous $\boldsymbol{\xi}$ and a $\boldsymbol{\Psi}$ covariance (*m*m*) matrix the covariances between any endogenous $\boldsymbol{\eta}$ (see, MacKinnon, 2008, p. 177). Equation 11 can be expanded to include the above $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ covariance matrices plus a $\boldsymbol{\Psi}$ covariance matrix for the correlation of error terms. Thus, the sample-based equivalent of the above $\boldsymbol{\Sigma}_y$ population matrix would be:

$$\boldsymbol{\Sigma}_y = \boldsymbol{\Lambda}\boldsymbol{\Psi}_{\eta\eta}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\boldsymbol{\Phi}_{\xi\xi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}\boldsymbol{\Psi}_{\eta\xi}\boldsymbol{\Theta}' \quad (12)$$

where $\boldsymbol{\Psi}_{\eta\eta}$ represents covariance of the *m*m* endogenous factors (the **B** matrix), the $\boldsymbol{\Phi}_{\xi\xi}$ represents the covariance of the *n*n* exogenous factors (the $\boldsymbol{\Gamma}$ matrix) and $\boldsymbol{\Psi}_{\eta\xi}$ the correlation matrix of the indicator errors $\boldsymbol{\Theta}_x$ and $\boldsymbol{\Theta}_y$ .

### 4.2 Testable assumptions under the above autoregressive model

The autoregressive model in Figure 1 can test a number of fundamental measurement assumptions explicitly that cannot be directly addressed in regression designs. Among those is the assumption of autocorrelated errors. If repeated measures do not control for autocorrelated error, estimates of indicator loadings and the overall fit of the model are compromised (Fabrigar *et al.*, 1999). This is because all repeated measures of the same indicator carry some proportion of shared method variance (Cole and Maxwell, 2003) which needs to be extracted from indicator measurement error. In a similar vein, correlated errors between two theoretically related measures both of which are indicators of the same construct (like father's and mother's socioeconomic status) can be specified.

The above autoregressive model can also test hypotheses about longitudinal invariance explicitly. To compare the three occasions representing repeated measures of the latent construct in Figure 1, the latent construct must remain similar across time. If this condition is not met, any interpretation of change among occasions is misleading. It is not certain whether the observed change is due to true change or change due to sampling variability in respondents' perceptions or interpretations, or other reasons (Chan, 1998).

Testing hypotheses of equivalence in the structural part typically proceeds after tests of equivalence of the measurement parts have been completed and the required level of longitudinal invariance achieved. Autoregressive models in MACS analyses offer three important pieces of information as regards the structural model: First, how much each occasion depends on its prior one. A high standardized coefficient between occasions signifies high dependence and less change between occasions. In case of more than two occasions, tests of structural equivalence of the paths connecting occasions can be made. Second, the latent means among consecutive occasions can be compared. This indicates whether average between-individual differences in the first occasion increased or decreased in subsequent occasions. Differences in latent means across occasions can be directly estimated and their significance assessed under a MACS framework. Contrary to the observed means, latent means are considered unbiased (Millsap, 2011). Third, estimation of latent intercepts in the endogenous constructs suggests how much net change in the subsequent latent construct has resulted from a prior occasion over time if the effect of the prior occasion was zero. A significant positive or negative latent intercept suggests that the net change in the latent means between longitudinally comparable occasions has been significant.

### 4.3 Assessing moderation

A moderator is any variable Z that changes the *direction* or *strength* of a causal relationship between an exogenous variable X and a endogenous variable Y, without being part of the causal model (Winkel, Saegert and Evans, 2009). Moderation analyses therefore concern the 'when' or 'for whom' the relation between a independent and dependent variable changes direction or strength when the levels or categories of the moderator change (Frazier, Tix and Barron, 2004, p. 116; Holmbeck, 1997). Naturally therefore, moderators are of substantive interest in the social sciences. They offer information on when or in which environments the intervention is more likely to be effective (Kraemer *et al.*, 2002). This study tests the hypothesis of whether maternal ethnicity moderates the hypothesised change parameters (dependence paths) in young people's educational expectations between ages 14-16.

In a multiple regression framework, moderation is typically assessed as an interaction term between the moderator (Mo) and the exogenous variable (X) while controlling for the main effects of both the moderator and the exogenous variables. This relation is typically is expressed (Jaccard and Turrisi, 2003) as:

$$Y = i + c_1X + c_2Mo + c_3XMo + e_y \qquad (13)$$

where Y is the dependent variable and $i$ is the intercept. Factoring out X, equation 13 becomes:

$$Y = i + (c_1 + c_3\,Mo)X + c_2Mo + e_y \qquad (14)$$

which suggests that the effect of X on Y is a function of Mo, called *conditional* effect of X on Y (Hayes, 2012, p. 4). Thus, $c_1$ estimates the effect of X on Y when Mo = 0 and $c_3$ estimates how much the effect of X on Y changes as Mo changes by one unit (ibid). The regression models 13 and 14 can be extended to include multiple moderators (Hayes, 2012; 2013).

However, despite their conceptual clarity and computational ease (requiring no specialised SEM software) moderator regression models with observed variables suffer from all the major limitations of multiple regression models with observed variables referred to above regarding their inability to partition measurement error (Gu, Preacher and Ferrer, 2014; Muthén and Asparouhov, 2014). Typically, models like those expressed in equations 13 and 14 are also cross-sectional, thus allowing no time to function between the predictor and the outcome (Gollob and Reichardt, 1987). Yet, these models are treated as causal (Edwards and Lambert-Schurer, 2007). Further, when the moderator is a discrete categorical variable (e.g., gender or ethnicity), Aguinis (1995) demonstrated that the power to detect a moderation effect depends on the difference in the sample sizes of the subgroups of the moderator. If ethnicity is measured by several ethnic groups of unequal sample sizes, then the power to detect moderation effects is severely reduced in multiple regression models (Aguinis, 1995, p. 1148). However, perhaps their most important limitation has to do with the interpretation of the effect of the moderator on the model parameters. It is expected that interaction terms should be specified in a regression model after a good theoretical reason. However, regression models that include two-, three- or four-way interaction terms are unlikely to be always guided by theory (Jaccard and Wan, 1995). More importantly, many of these interaction terms will be statistically insignificant simply as a function of sampling variability (Aguinis, 1995) while their interpretation substantively problematic.

Instead of interaction terms, the present analysis applies a multigroup approach to test for moderation by maternal ethnicity. This approach has been advocated early on (Baron and Kenny, 1986; Jaccard and Wan, 1995) particularly when SEM methodologies can be applied (Little *et al.*, 2007; Muller, Judd and Yzerbyt, 2005). The advantages of this approach are multi-fold: First, each parameter of the measurement and structural model can be tested for moderation. Tests for moderation in a CFA framework have more statistical power because they are typically based on a greater number of fit indices. By contrast, moderated multiple regression models have less statistical power, particularly if there is sample inconsistency in the moderator categories (Aguinis, 1995; Aguinis, Petersen and Pierce, 1999). Second, moderation in a CFA framework can be studied on every structural relation. Thus, it becomes easy to see which relations out of a complex web of hypothesized relations are moderated and which are not. Asymptotic estimates can be bootstrapped for greater confidence in their interpretation. Third, every *structural* effect can be compared across groups after cross-group measurement invariance has been established. Thus, obtained cross-group differences in structural estimates can be more reliably attributed to group membership characteristics (i.e., maternal ethnicity). Fourth and most importantly, the potentially moderated structural direct and indirect effect are adjusted for measurement error, prior occasions and method and trait variance.

When the MACS analysis involves more than one group, representing different subgroups of the population as is the case here, cross-group measurement invariance needs to be established in addition to longitudinal invariance. Typically, tests of longitudinal measurement invariance precede tests of cross-group measurement invariance but the order can be reversed (Millsap, 2011). The latent constructs must share a minimum level of cross-group measurement invariance so that the structural parts of the models can be compared across groups as defined by the hypothesised moderators. Significant cross-group differences in latent means and intercepts suggest the presence of moderation of these structural parameters by group membership. Cross-group comparisons of structural path estimates are therefore central in this analysis and are supplemented by the cross-group comparisons of means and intercepts. As in the case of latent means and intercepts, statistically significant differences between ethnic groups in structural paths signify moderation by ethnic group membership. Autoregressive models like the one shown in Figure 1 are capable therefore of gauging moderation in both their measurement (indicator loadings, intercepts and errors) and structural (latent means, intercepts and structural paths) models (Byrne, 2010).

## 6. Findings

### 6.1 Meeting the required measurement assumptions

Measurement invariance/equivalence (MI/E) expresses the degree to which comparisons between latent constructs are possible by assessing whether latent constructs remain equivalent over time or across groups of different membership (Bagozzi and Yi, 1988; Dimitrov, 2010). Invariance tests are inherently therefore tests of the potential effect of a *moderator*, either *time* (in longitudinal invariance) or *group membership k* (in cross-group invariance) (Palich, Horn and Griffeth, 1995; Riordan and Vandenberg, 1994). Unless an acceptable degree of longitudinal or cross-group construct equivalence is established, any comparisons among constructs are misleading. Without appropriate MI/E tests, it is impossible to know whether the observed change over time was due to true development; moderation by group membership; or because the construct was perceived or interpreted differently. Acceptable levels of MI/E permit longitudinal and cross-group comparisons of structural estimates including latent means and intercepts (Widaman, Ferrer and Conger, 2010). For the purposes of the present analysis, invariance tests establish whether all latent constructs of pupils' educational expectations share particular psychometric properties necessary to their comparison. At a minimum, latent constructs and structural relations among them are comparable only if the constructs share an identical form (configural invariance); their indicators exhibit the same relationship to the latent construct (metric invariance) and share the same origin (scalar invariance) (Brown, 2006; Meredith, 1993).
The chi-square difference test ($\Delta\chi^2$) is crucial in invariance tests. The chi-square value and the associated *p* value can be interpreted unambiguously because its distribution is known (Arbuckle, 2009). However, because chi-square depends on sample size, it may also lead to the rejection of the null hypothesis too often, thus increasing Type I error rates or 'false positives' (i.e., showing significant differences while in reality there are none). In SEM, a nonsignificant chi-square is desired to indicate no difference between model and data and thus, good fit to data (Bentler and Bonett, 1980, p. 591). However, with large sample sizes the test will show that the data are significantly different even though the difference is 'so very slight as to be negligible or unimportant on other criteria (Gulliksen and Tukey, 1958, p. 95-96). To adjust for this possibility, methodologists have advised that the chi-square should be interpreted in conjunction with other fit indices (Arbuckle, 2011; Bollen, 1989b; Bollen and Long, 1993; Hu and Bentler, 1993; Widaman, Ferrer and Conger, 2010). Tanaka (1993, p. 32) proposed a set of alternative fit indices indicating when *not* to reject the null hypothesis solely on the basis of a significant model chi-square or the chi-square difference value.

In addition to the model chi-square, I also report the *normed fit index* (NFI) (Bentler and Bonett, 1980) or *delta 1* (Bollen, 1989b). Values ≥ 0.90 indicate good fit and ≥ 0.95 very good fit (Bollen, 1989a); The *Relative fit index* (RFI) or *rho 1* (Bollen, 1986) ranges from 0 – 1 with values ≥ 0.90 indicating good fit and ≥ 0.95 very good fit (Arbuckle, 2011). The *incremental fit index* (IFI) (Bollen, 1989b) also referred to as *delta 2* (Bollen, 1989b; 1990) adjusts for sample size dependency of NFI (Tanaka, 1993, p. 36). The *Tucker-Lewis coefficient* (TLI), also known as *rho 2* (Bollen, 1989b) or *non-normed fit index* (NNFI) compares the fit of the target model to that of the null model. Its advantage is that it is not sample-dependent (Tanaka, 1993, p. 32). Typically it ranges from 0-1 but in very-well fitting models it can slightly exceed 1.00. (5) The *comparative fit index* (CFI) (Bentler, 1990) adjusts for the noncentrality parameter, taking into consideration the non-normality of the chi-square distribution (Marsh, Balla and McDonald, 1988). CFI is identical to the *relative noncentrality index* (RNI) (McDonald and Marsh, 1990) but it is normed so that it has a range of 0-1. Values $\geq 0.90$ indicate good fit and $\geq 0.95$ very good fit (Vandenberg and Lance, 2000). Finally, *Root mean square* (RMS) (Steiger and Lind, 1980) or *root mean square error of approximation* (RMSEA) (Browne and Cudeck, 1993) is also a noncentrality-based index but also adjusts for model complexity and sample size. It is particularly sensitive to misspecified factor loadings (Hu and Bentler, 1998; Hu and Bentler, 1999) so it is good to use for metric invariance tests. It expresses the error of the target model to approximate the true model in the population (Jöreskog and Sörbom, 1996, p. 124). RMSEA produces a 90% two-tailed confidence interval with a lower bound of 0.0 and an upper bound of $+\infty$. RMSEA values $\leq 0.05$ indicate close fit while values $\leq 0.08$ indicate a reasonable error of approximation (Browne and Cudeck, 1993). The PCLOSE is also reported in connection to the RMSEA. It expresses a *p* value for testing the hypothesis that the true value of the RMSEA is $\leq 0.05$. It ranges from 0 to 1.00.

However, the distributions of the above indices are unknown (Arbuckle, 2009). Therefore, differences in values of any index will also have unknown distributions. Cheung and Rensvold (1999) studied the behaviour of TLI, RMSEA and CFI under varying conditions of measurement invariance, sample size discrepancy in multigroup solutions and model specification. They concluded that all three indices were superior to chi-square in terms of Type I error rates. They suggested that a change in CFI of -0.01 or less indicated that the null hypothesis of invariance should not be rejected. Other studies reported that a change in TLI of $\leq 0.02$ (McGaw and Jöreskog, 1971) or $\leq 0.05$ (Little, 1997) was negligible and the invariance hypothesis should not be rejected. More recent simulation studies have confirmed that CFI and RMSEA were robust to varying sample sizes, under multiple conditions of factorial invariance (Cheung and Rensvold, 2002). A CFI change ($\Delta$CFI) of $\leq 0.01$ and a RMSEA change ($\Delta$RMSEA) of $\leq 0.016$ (see, Cheung and Rensvold, 2002, Table 4, p. 245) were reasonable indications that the hypothesis of invariance should not be rejected even if $\Delta\chi^2$ was significant. If noninvariance cannot be supported by theory or previous research, then greater reliance on the alternative fit indices is recommended (Raykov, 2004; Widaman, Ferrer and Conger, 2010).

Table 6 reports the model fit for the model in Figure 1 with configural, metric and scalar longitudinal invariance constraints in place. The table reports the chi-square value ($\chi^2$), degrees of freedom (df), significance (p) and the discrepancy/df ratio ($\hat{C}$/d) as well as normed fit index (NFI); relative fit index (RFI); incremental fit index (IFI); Tucker-Lewis index (TLI); comparative fit index (CFI); the root mean square error of determination (RMSEA); its lower (LO) and upper (HI) bounds and the probability that the RMSEA is $\leq 0.05$ in the population (PCLOSE). As a yardstick, I also report the cut-off points above which indicate very good to excellent fit at the bottom of Table 6 and the sample sizes next to each ethnicity group.

Table 6: Fit to data of the model of pupils' educational expectations for each ethnic group

| | $\chi^2$ | df | p | *C*/d | NFI | RFI | IFI | TLI | CFI | RMSEA | LO | HI | PCLOSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| White (n=1000) | 15.9 | 10 | ns | 1.586 | 0.996 | 0.993 | 0.998 | 0.998 | 0.998 | 0.024 | 0.000 | 0.046 | 0.979 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Indian (n=751) | 6.0 | 9 | ns | 0.661 | 0.997 | 0.994 | 1.002 | 1.003 | 1.000 | 0.000 | 0.000 | 0.030 | 1.000 |
| Pakistani (n=642) | 14.6 | 9 | ns | 1.621 | 0.989 | 0.982 | 0.996 | 0.993 | 0.996 | 0.031 | 0.000 | 0.059 | 0.850 |
| Bangladeshi (n=487) | 13.5 | 12 | ns | 1.123 | 0.987 | 0.984 | 0.999 | 0.998 | 0.999 | 0.016 | 0.000 | 0.05 | 0.947 |
| BCaribbean (n=324) | 20.3 | 10 | 0.0 | 2.031 | 0.969 | 0.954 | 0.984 | 0.976 | 0.984 | 0.056 | 0.019 | 0.092 | 0.338 |

Note: $\chi^2$ =chi-square; df=degrees of freedom; p=significance; *C*/d=ratio of discrepancy to df (<2.00); NFI=normed fit index (>0.95); RFI=relative fit index (>0.95); IFI=incremental fit index (>0.95); TLI=Tucker-Lewis index (>0.95); CFI=comparative fit index (>0.95); RMSEA=Root mean square error or approximation (<0.05); LO=lower bound of the RMSEA (0.0); HI=upper bound of the RMSEA (<.0.05); PCLOSE=probability that **the RMSEA is ≤ 0.05 in the population (1.000).**

The model of adolescent educational expectations exhibits excellent fit to data. In almost all ethnic groups, the chi-square was statistically nonsignificant, even under the constraints of longitudinal equivalence of indicator loadings (metric invariance) and intercepts (scalar invariance). In the case of Black Caribbean pupils, the chi-square was significant but the rest of the fit indices still indicated excellent fit. Regarding the actual parameter estimates for the measurement part, Table 7 reports the ML estimates for the factor loadings while Table 8 reports the indicator intercepts for each ethnic group.

Table 7: Factor loadings of the model of adolescent educational expectations for each ethnic group

| | White | | | Indian | | | Pakistani | | | Bangladeshi | | | BCaribbean | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | b | SE | β | b | SE | β | b | SE | β | b | SE | β | b | SE | β |
| $\lambda_{11}$ *How likely to apply to university* w1 | 0.603 | 0.015 | 0.743 | 0.673 | 0.028 | 0.634 | 0.661 | 0.032 | 0.682 | 0.698 | 0.034 | 0.689 | 0.626 | 0.040 | 0.649 |
| $\lambda_{21}$ *How likely to get in university if apply* w1 | 1.000 | | 0.903 | 1.000 | | 0.850 | 1.000 | | 0.869 | 1.000 | | 0.864 | 1.000 | | 0.828 |
| $\lambda_{12}$ *How likely to apply to university* w2 | 0.603 | 0.015 | 0.756 | 0.673 | 0.028 | 0.677 | 0.661 | 0.032 | 0.678 | 0.698 | 0.034 | 0.695 | 0.626 | 0.040 | 0.700 |
| $\lambda_{22}$ *How likely to get in university if apply* w2 | 1.000 | | 0.923 | 1.000 | | 0.871 | 1.000 | | 0.844 | 1.000 | | 0.875 | 1.000 | | 0.823 |
| $\lambda_{13}$ *How likely to apply to university* w3 | 0.603 | 0.015 | 0.781 | 0.673 | 0.028 | 0.693 | 0.661 | 0.032 | 0.739 | 0.698 | 0.034 | 0.703 | 0.626 | 0.040 | 0.660 |
| $\lambda_{33}$ *How likely to get in university if apply* w3 | 1.000 | | 0.924 | 1.000 | | 0.918 | 1.000 | | 0.880 | 1.000 | | 0.878 | 1.000 | | 0.888 |

Note: b=unstandardized loading; **β**=standardized loading; SE=standard error. 1.000 under (b) refers to loadings which were fixed to unity for identification purposes. Like indicators were fixed to unity as a requirement of configural longitudinal and cross-group invariance

Table 8:ML estimates for indicator intercepts for each ethnic group

| | White | | Indian | | Pakistani | | Bangladeshi | | BCarib | |
|---|---|---|---|---|---|---|---|---|---|---|
| | τ | SE | τ | SE | τ | SE | τ | SE | τ | SE |
| $\tau_1$ *How likely to apply to university* w1 | 2.944 | 0.019 | 3.276 | 0.017 | 3.204 | 0.018 | 3.146 | 0.021 | 3.132 | 0.026 |
| $\tau_2$ *How likely to get in university if apply* w1 | 2.980 | 0.027 | 3.513 | 0.020 | 3.307 | 0.022 | 3.238 | 0.026 | 3.216 | 0.035 |
| $\tau_3$ *How likely to apply to university* w2 | 2.944 | 0.019 | 3.276 | 0.017 | 3.204 | 0.018 | 3.146 | 0.021 | 3.132 | 0.026 |
| $\tau_4$ *How likely to get in university if apply* w2 | 2.980 | 0.027 | 3.513 | 0.020 | 3.307 | 0.022 | 3.238 | 0.026 | 3.216 | 0.035 |
| $\tau_5$ *How likely to apply to university* w3 | 2.944 | 0.019 | 3.276 | 0.017 | 3.204 | 0.018 | 3.146 | 0.021 | 3.132 | 0.026 |
| $\tau_6$ *How likely to get in university if apply* w3 | 2.980 | 0.027 | 3.513 | 0.020 | 3.307 | 0.022 | 3.238 | 0.026 | 3.216 | 0.035 |

Tables 7 and 8 suggest that all measurement models were longitudinally consistent, i.e., as suggested by the unstandardised (**b**) and standardised (**β**) coefficients, the loadings (**λ**) of like indicators across groups had standardized estimates of roughly equal magnitudes. The fact that the chi-square was insignificant in all but the Black Caribbean group also suggests that configural, metric and scalar longitudinal invariance has been achieved. However, this is confirmed in the tests of longitudinal invariance reported in Table 9 below. For each group, the unconstrained baseline model is sequentially compared to a model with only equality of loadings imposed (labelled 'metric' in the Table). This model is then compared to a third model in which both indicator loadings as well as intercepts are constrained to longitudinal equality (labelled ‘scalar’ in the Table).

Table 9: Tests of longitudinal invariance of the model of adolescent educational expectations for each

| | $\chi^2$ | df | p | $\Delta\chi^2$ | df | p | CFI | ΔCFI | RMSEA | ΔRMSEA |
|---|---|---|---|---|---|---|---|---|---|---|
| White (n=1000) | | | | | | | | | | |
| baseline | 7.6 | 4 | ns | - | - | - | 0.999 | - | 0.030 | - |
| metric | 7.6 | 6 | ns | 0.0 | 2 | ns | 1.000 | 0.001 | 0.017 | -0.013 |
| scalar | 15.9 | 10 | ns | 8.3 | 4 | ns | 0.998 | -0.002 | 0.024 | 0.007 |
| Indian (n=751) | | | | | | | | | | |
| baseline | 3.8 | 3 | ns | - | - | - | 1.000 | - | 0.018 | - |
| metric | 3.9 | 5 | ns | 0.1 | 2 | ns | 1.000 | 0.00 | 0.000 | -0.018 |
| scalar | 6.0 | 9 | ns | 2.1 | 4 | ns | 1.000 | 0.00 | 0.000 | 0.000 |
| Pakistani (n=642) | | | | | | | | | | |
| baseline | 1.7 | 3 | ns | - | - | - | 1.000 | - | 0.000 | - |
| metric | 3.9 | 5 | ns | 2.2 | 2 | ns | 1.000 | 0.00 | 0.000 | 0.000 |
| scalar | 14.6 | 9 | ns | 10.7 | 4 | 0.05 | 0.996 | -0.004 | 0.031 | 0.031 |
| Bangladeshi (n=487) | | | | | | | | | | |
| baseline | 5.7 | 6 | ns | - | - | - | 1.000 | - | 0.000 | - |
| metric | 6.8 | 8 | ns | 1.1 | 2 | ns | 1.000 | 0.00 | 0.000 | 0.000 |
| scalar | 13.5 | 12 | ns | 6.7 | 4 | ns | 0.999 | -0.001 | 0.016 | 0.016 |
| Black Caribbean (n=324) | | | | | | | | | | |
| baseline | 10.8 | 5 | ns | - | - | - | 0.991 | - | 0.060 | - |
| metric | 13.3 | 7 | ns | 2.5 | 2 | ns | 0.990 | -0.001 | 0.053 | -0.007 |
| scalar | 20.3 | 10 | 0.0 | 7.0 | 3 | 0.05 | 0.984 | -0.006 | 0.056 | 0.003 |

Note: $\chi^2$ =chi-square; df=degrees of freedom; p=significance; $\Delta\chi^2$= chi-square difference; CFI=comparative fit index (>0.95); RMSEA=Root mean square error or approximation (<0.05); **ΔCFI=change in CFI (≤** -0.01); **ΔRMSEA=change in RMSEA (≥ 0.016); ‘p.metric’ or ‘p.scalar’=partial metric/partial scalar**

Table 9 confirms that the model of adolescent educational expectations was longitudinally invariant across all ethnic groups. The model exhibited full metric and full scalar invariance in all groups. This suggests that the latent construct measuring educational expectations was similar across all occasions in all groups. With configural, full metric and full scalar longitudinal invariance achieved, the structural parameters, latent means and latent intercepts can be compared over time. These structural parameters appear in Table 11. Bias-corrected two-tailed significance (p) is reported based on the percentile method using 1000-sample Monte Carlo parametric bootstrap.

Table 11: The structural part of the model of adolescent educational expectations across the five ethnic groups

| | White | | | Indian | | | Pakistani | | | Bangladeshi | | | BCaribbean | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | b | p | **β** | b | p | **β** | b | p | **β** | b | p | **β** | b | p | **β** |
| $p_{21}$ | 0.922 | 0.002 | 0.819 | 0.787 | 0.002 | 0.690 | 0.541 | 0.003 | 0.527 | 0.711 | 0.002 | 0.655 | 0.733 | 0.002 | 0.703 |
| $p_{32}$ | 0.653 | 0.003 | 0.645 | 0.651 | 0.002 | 0.644 | 0.728 | 0.002 | 0.672 | 0..579 | 0.002 | 0.605 | 0.596 | 0.002 | 0.590 |

Note: b=unstandardized weight; **β**=standardized weight; p=bias-corrected percentile method, 2-tailed significance.

Based on the above tests, the standardised structural parameters of Table 11 (in bold) can be compared over time in each group. If tests of structural invariance over time are significant, they will suggest that the above structural parameters change significantly over time (i.e., they are moderated by time). Estimates of higher magnitude in the time-dependence structural paths ($p_{21}$, $p_{32}$) suggest *less* change (or greater stability) from one occasion to the next. Lower magnitudes suggest *higher* change (or less stability) from one occasion to the next because each occasion depends less on its prior measurement. Pending confirmation by longitudinal invariance tests that follow, Table 11 suggests that the white, Black Caribbean and Pakistani pupils exhibit greater change in their expectations as paths $p_{21}$ and $p_{32}$ are quite different in magnitude n these groups. By contrast, Indian and Bangladeshi pupils show greater stability (or less change) in their expectations during the same period.

### 6.2 Assessing moderated mediation in pupils' expectations for university study

Paths $p_{21}$ and $p_{32}$ reported in Table 11 represent *direct* effects. Path $p_{21}$ is an estimate of the direct effect of occasion 1 (expectations at age 14) on occasion 2 (expectations at age 15) and $p_{32}$ is an estimate of the direct effect of occasion 2 on occasion 3 (expectations at age 16), controlling for the influence of occasion 1. Direct effects can be interpreted as regression coefficients in multiple regression models with observed variables. However, occasion 3 is also affected indirectly by occasion 1 via occasion 2. In contrast to multiple regression models, analyses under the present psychometric framework can decompose a total effect into direct and indirect. In this case, the total effect of expectations at age 14 on expectations at age 16 can be decomposed into direct ($p_{31}$) and indirect effect which is calculated as the product of $p_{21}$ and $p_{32}$ ($p_{21}p_{32}$). Comparison of the indirect effect across ethnic groups gives a measure of the extent to which mediation of the effect of earlier expectations at age 14 on expectations at age 16 via expectations at age 15 is moderated by maternal ethnicity. This indirect effect can be bootstrapped for an estimate of its statistical significance. Table 12 reports the standardised (β) and unstandardised (b) direct, indirect and total effect of expectations at age 14 on expectations at age 16. Significance for Table 12 is based on the bootstrapped bias-corrected p value for each sample of mothers (based on 1000 bootstrapped samples).

Table 12: Standardised (**β**) and unstandardised (b) direct, indirect and total effect of expectations at age 14 on expectations at age 16

| Effect | White | | | Indian | | | Pakistani | | | Bangladeshi | | | Black Caribbean | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **β** | b | p | **β** | b | p | **β** | b | p | **β** | b | p | **β** | b | p |
| Indirect | .528 | .602 | (.001) | .444 | .513 | (.002) | .354 | .394 | (.002) | .396 | .412 | (.002) | .415 | (.001) | (.004) |
| Direct | .223 | .254 | (.002) | .108 | .124 | (.095) | .100 | .111 | (.049) | .110 | .114 | (.112) | .160 | (.168) | (.004) |
| Total | .751 | .856 | (.003) | .552 | .637 | (.002) | .454 | .505 | (.002) | .506 | .526 | (.002) | .575 | (.605) | (.002) |

Tables 11 and 12 suggest that expectations at ages 15 and 16 are primarily driven directly by prior expectations at age 14 and 15. Prior expectations at age 14 in all groups except the white have an insignificant direct influence on later expectations at age 16. Most importantly however, expectations at age 14 exert a very significant *indirect* effect on expectations at age 16 over and above the immediate direct effects of expectations at age 15. The largest magnitude of this indirect effect is found in the white pupils and the smallest in the Pakistani pupils. The above decomposition shows that adolescent expectations at age 16 are shaped by very significant direct and indirect influences from prior expectations at ages 14 and 15. We need to know both types of effects to understand how expectations change longitudinally.  I will complete the description of those effects when I place them in proper context when the actual level of the latent construct representing expectations is actually known.

**(a) Do adolescent educational expectations change over ages 14-16?**

Table 11 suggests that white and Pakistani pupils' educational expectations were less stable (less stationary) over ages 14 to 16 compared to those of their Black Caribbean, Indian and Bangladeshi peers. But is this observed change in the magnitudes of the dependence paths statistically significant? If this is the case, and given the high levels of longitudinal invariance achieved, we shall have a good indication of true change between pupils' expectations at ages 14 and 15 and between ages 15 and 16. Table 13 presents the results of longitudinal structural invariance for every minority ethnic group. The null hypothesis of longitudinal structural invariance ($H_0$: $p_{21} = p_{32}$) is tested in the second row and the chi-square difference test ($\Delta\chi2$) is reported across this row. The decision to reject (R) or not reject (NR) the null hypothesis of longitudinal structural invariance is reported in the last column. This decision is based on both the significance of the chi-square difference tests as well as the reported change in the CFI ($\Delta$CFI) and the RMSEA ($\Delta$RMSEA).

Table 13: Tests of longitudinal invariance (stationarity) of the dependence paths P21 and p32 in each ethnic group

| Hypothesis | $\chi2$ | df | p | $\Delta\chi2$ | df | p | CFI | $\Delta$CFI | RMSEA | $\Delta$RMSEA | Decision |
|---|---|---|---|---|---|---|---|---|---|---|---|
| White | | | | | | | | | | | |
| $p_{21} \neq p_{32}$ | 15.1 | 9 | .089 | - | - | - | .998 | - | .026 | - | - |
| $p_{21} = p_{32}$ | 29.4 | 10 | .001 | 14.3 | 1 | .00 | .995 | -.003 | .044 | .018 | R |
| Indian | | | | | | | | | | | |
| $p_{21} \neq p_{32}$ | 2.3 | 7 | ns | - | - | - | 1.000 | - | .000 | - | - |
| $p_{21} = p_{32}$ | 5.1 | 8 | .744 | 2.8 | 1 | ns | 1.000 | .00 | .000 | .00 | NR |
| Pakistani | | | | | | | | | | | |
| $p_{21} \neq p_{32}$ | 28.1 | 9 | .00 | - | - | - | .986 | - | .031 | - | - |
| $p_{21} = p_{32}$ | 34.1 | 10 | .00 | 6 | 1 | .014 | .982 | -.014 | .061 | .03 | R |
| Bangladeshi | | | | | | | | | | | |
| $p_{21} \neq p_{32}$ | 11.2 | 10 | ns | - | - | - | .999 | - | .016 | - | - |
| $p_{21} = p_{32}$ | 13.2 | 11 | ns | 2 | 1 | ns | .998 | .001 | .020 | .004 | NR |
| Black Caribbean | | | | | | | | | | | |
| $p_{21} \neq p_{32}$ | 24.5 | 10 | .006 | - | - | - | .978 | - | .067 | - | - |
| $p_{21} = p_{32}$ | 25.7 | 11 | .007 | 1.2 | 1 | ns | .977 | .001 | .064 | -.003 | NR |

Note: $\chi^2$ =chi-square; df=degrees of freedom; p=significance; $\Delta\chi^2$= chi-square difference; CFI=comparative fit index (>0.95); $\Delta$CFI= CFI difference test; RMSEA=Root mean square error or approximation (<0.05); $\Delta$RMSEA=RMSEA difference test; R=reject $H_o$; NR=fail to reject the $H_o$;

Based on the information suggested by Table 13, the null hypothesis of structural invariance in the model parameters could be rejected only in the case of white and Pakistani pupils. This means that in the case of the white and Pakistani pupils, change in expectations between ages 14 to 16 and between 15 to16 was significantly different. But, while white pupils' expectations tended to become less stable between ages 15 to 16, the opposite was the case in the Pakistani group where expectations became more stable. Table 11 shows some change in the magnitude of these structural parameters (in bold), in the case of Indian, Bangladeshi and Black Caribbean pupils as well, indicating a more universal tendency for expectations to change more from age 15 to age 16. However, this is not confirmed by the significance of the differences in the structural parameters $p_{21}$ and $p_{32}$ of these groups with their measurement model held to 'strong' (Meredith, 1993) measurement invariance. Substantively this evidence suggests that the amount of change (or stability) in pupils' educational expectations between ages 14 to 15 and between ages 15 and 16 was statistically equivalent in all groups except the white and Pakistani. For all groups in other words, except the white and Pakistani, pupils' expectations between ages 14 to 15 changed as much as they did between ages 15 to 16. In both the white and the Pakistani group however, expectations appear to change significantly less between ages 14 to 15 and more between ages 15 to 16. I will revisit this point later.

**(b) Is longitudinal change in educational expectations moderated by maternal ethnicity?**

However, the information suggested by Tables 11 and 12 is inherently limited. We do not know whether change from one occasion to the next represents stability at a high, medium or low level of the latent dimension. High stability in *low* expectations over time tells quite a different story from low stability in *high* expectations, for example. Similarly, strong indirect effects of low expectations at age 14 on expectations at age 16 have quite different substantive implications from strong indirect effects of high expectations. Change in latent means and intercepts over time offers this missing information. If appropriate levels of cross-group measurement invariance are established, structural estimates become comparable across groups. As a result, we can assess whether structural parameter estimates, factor means and factor intercepts were moderated by maternal ethnicity.

In testing for longitudinal measurement invariance, the same criteria for the rejection of the null hypothesis were also applied as in the testing for cross-group measurement invariance. Identification of noninvariant loadings or intercepts was guided by modification indices (MI). MIs represented the expected change in the model chi-square based on the Lagrange multiplier (Arbuckle, 2009) if a particular constraint was removed. Following Byrne *et al*., (Byrne, Shavelson and Muthén) and Byrne (2004), the highest MI was freely estimated first. Standard procedure in cross-group invariance testing (Stacy, MacKinnon and Pentz, 1993; Tyson, 2004; Vandenberg and Lance, 2000; Wicherts, Dolan and Hessen, 2005) suggested that each occasion should be tested for metric and scalar cross-group invariance separately.

Table 14 below describes the tests for cross-group measurement invariance. ‘Baseline’ refers to the unrestricted model. ‘Full metric’ and ‘full scalar’ refer to cases where full metric and full scalar invariance was achieved. In cases where partial scalar invariance was achieved ‘free’ refers to the freely-estimated intercepts. Group membership of those freely-estimated items is identified by W=white; I=Indian; P=Pakistani, B=Bangladeshi and BC=Black Caribbean. Once a level of cross-group invariance was established in one occasion, the constraints placed on items that were shown to be cross-group invariant were retained when testing the level of invariance of the next occasion. When tests of metric invariance for all occasions in a model were complete, the first occasion was again tested for cross-group scalar invariance, and so on. In Table 13, levels of cross group invariance are separated by bold horizontal lines in each group.

Free estimation of noninvariant items across groups resulted in a stepwise improvement of overall model fit. This improvement can be followed by the *positive* values of CFI difference tests (ΔCFI) and the *negative* values of the RMSEA difference tests (ΔRMSEA) suggesting increase of the model CFI and decrease of the RMSEA, both of which signify improvement of fit (Dimitrov, 2010). When tests of scalar invariance were complete, the model that achieved the highest level of metric invariance (labelled ‘metric occasion 3’) was compared to the model that achieved the highest level of scalar invariance (labelled ‘scalar occasion 3’). The chi-square difference between the two models showed whether the more constrained model with scalar cross-group invariance constraints deteriorated the model fit achieved with only metric cross-group invariance constraints. This $\Delta\chi^2$ test appears in the last row in Table 13. The more constrained model never deteriorated overall model fit so as to exceed the recommended cut-off points for the CFA (ΔCFA < -0.01) and the RMSEA (ΔRMSEA > 0.016) (Cheung and

Rensvold, 2002). In all cases, the fit of the final multigroup solution with metric and scalar cross-group invariance constraints in place (shown in bold in Table 14) was excellent.

Table 14: Tests of cross-group measurement invariance

| Level of invariance tested in multigroup solution | $\chi^2$ | df | p | $\Delta\chi^2$ | df | p | CFI | ΔCFI | RMSEA | ΔRMSEA |
|---|---|---|---|---|---|---|---|---|---|---|
| Baseline (unconstrained) | 28.3 | 19 | ns | - | - | - | 0.999 | - | 0.012 | - |
| Full metric occasion 1 | 33.6 | 23 | ns | 5.3 | 4 | ns | 0.999 | 0.000 | 0.012 | 0.000 |
| Full metric occasion 2 | 39.8 | 27 | ns | 6.2 | 4 | ns | 0.998 | -0.001 | 0.012 | 0.000 |
| Full metric occasion 3 | 50.3 | 31 | 0.0 | 10.5 | 4 | 0.0 | 0.998 | 0.000 | 0.014 | 0.002 |
| Partial scalar occasion 1 | 244.0 | 39 | 0.0 | 198.7 | 8 | 0.0 | 0.976 | -0.022 | 0.041 | 0.027 |
| Free: W1heposs I | 169.6 | 38 | 0.0 | 74.4 | 1 | 0.0 | 0.984 | 0.008 | 0.033 | -0.008 |
| Free: W1hlike W | 128.1 | 37 | 0.0 | 41.5 | 1 | 0.0 | 0.989 | 0.005 | 0.028 | -0.004 |
| Full scalar: occasion 2 | 147.2 | 43 | 0.0 | 19.1 | 5 | 0.0 | 0.998 | 0.009 | 0.028 | 0.000 |
| Full scalar: occasion 3 | 177.5 | 53 | 0.0 | 30.3 | 10 | 0.0 | 0.994 | -0.004 | 0.029 | 0.001 |
| Metric occasion 3 – scalar occasion 3 | | | | 127.2 | 22 | 0.0 | | -0.004 | | 0.015 |

Note: $\chi^2$ =chi-square; df=degrees of freedom; p=significance; $\Delta\chi^2$= chi-square difference; CFI=comparative fit index (>0.95); RMSEA=Root mean square error or approximation (<0.05); **ΔCFI=change in CFI (≤** -0.01); **ΔRMSEA=change in RMSEA (≥ 0.016). W=white; I=Indian; P=Pakistani, B=Bangladeshi and BC=Black** Caribbean

Configural cross-group invariance was established as the unconstrained baseline model of pupils' educational expectations in the multi-group solution had excellent fit. Full metric cross-group invariance was also established: Model fit did not deteriorate significantly with loading invariance constraints in place based on the change of CFI (ΔCFI column, Table 14) and RMSEA (ΔRMSEA column, Table 14). This level of invariance permits the cross-group comparison of the structural estimates that are shown in Table 11. With scalar measurement invariance established, indicator intercepts become directly comparable. The estimated indicator intercepts across all ethnic minority groups appear in Table 8 above.

The comparison of indicator intercepts suggests that the intercepts of both indicators ('*how likely to apply to university*' and '*how likely to get in if apply*') in the white pupils are the lowest across ages 14, 15 and 16, even compared to those of their Black Caribbean peers. As Table 13 reports, two of these intercepts in the white and the Indian groups were statistically cross-group non-invariant in occasion 1 representing pupils' expectations at age 14. Thus, there was statistical evidence of *differential item functioning* (DIF) in white pupils and the Indian pupils at age 14 ($\boldsymbol{\tau}_2$, $\boldsymbol{\tau}_1$). The intercepts of the white pupils were lower than those of their Indian peers suggesting different perceptions regarding their likelihood to enter university after year 11. This is hardly surprising considering that Indian pupils had the highest while white pupils the lowest proportions among those who regarded both their university application and their acceptance if apply 'very likely'. This gives strong support to the hypothesis that maternal ethnicity, to the extent it reflected different cultures, values and perceptions, moderated these responses. However, systematic examination of latent means and intercepts will confirm this fact.

**(c) Are cross-group differences in structural estimates moderated by maternal ethnicity?**

We are now in a position to bring together the results of the above tests. Regarding the longitudinal model of pupils' expectations between ages 14 to 16, it was shown that it supported both longitudinal and cross-group metric and scalar measurement invariance. This permitted the direct comparison of the structural parameter estimates both longitudinally (within each group) and across ethnic groups. So far, it was shown that *time* moderated longitudinal change only in white and Pakistani pupils' expectations between ages 14 to 16. It was only in those two groups where change in expectations between ages 14 to 15 and between age 15 to 16 was significantly different. The achieved level of cross-group invariance however further permits systematic tests of the hypothesis that in addition to time, maternal ethnicity also moderates cross-group differences in the same structural parameters.

This time, the question becomes whether the dependence paths $p_{21}$, and $p_{32}$, in the model of educational expectations, were statistically equivalent across groups. If those paths are found to be noninvariant across ethnicity groups, this noninvariance would suggest evidence of moderation by maternal ethnicity. A number of $c = [k*(k-1)/2]$ pairwise comparisons were conducted, where *k* represented the number of groups in the analysis. Since there were 5 groups in total in the analysis, there were a total of $c = [5(5-1)/2] = 10$ comparisons. Because each group was sequentially compared to all others, there was a higher likelihood of getting a result that would be significant at the $\alpha = 0.05$ level purely by chance, thus increasing Type I error. For this reason, a Bonferroni correction adjusted for the α level of the number of pairwise comparisons representing the *family-wise* Type I error rate given by

$$\alpha_{FW} = 1 - (1- \alpha)^{c} \qquad (15)$$

where c = number of pairwise comparisons as described in Bland and Altman (1995).

There is a debate as to the usefulness of this adjustment. While the Bonferroni adjustment decreases Type I error rates, it also increases Type II error rates, making it more likely to fail to identify significant differences (Perneger, 1998; Rothman, 1990). However, the risk of increasing Type I error (false positives) was much more important in this analysis because it would mean that moderation by maternal ethnicity was identified while in reality there was none. Thus, the Bonferroni correction was implemented making the α levels for the identification of such potential moderation effects *less* sensitive. This was done by dividing the $\alpha = 0.05$ level by the number of comparisons ($\alpha_{FW} = \alpha / c$) involving the same group. Since the same group was involved in four ($k$-1) comparisons, the α level was decreased from 0.05 to 0.0125 (Bhandari *et al.*, 2003). Thus, the hypothesis of equality between two structural parameters was rejected at $\alpha \leq 0.0125$. Although the minimum requirement for the comparison of structural estimates is *metric* cross-group invariance (Byrne, Shavelson and Muthén, 1989; Vandenberg and Lance, 2000), I compared models whose measurement part was constrained to both metric and scalar invariance.

Before systematic pairwise tests commenced, I conducted an omnibus test of cross-group structural invariance, testing the hypothesis that all $p_{21}$ paths ($H_o$: $p_{21k} = p_{21}$) and all $p_{32}$ paths ($H_o$: $p_{32k} = p_{32}$) where $k$ = group membership, were cross-group invariant. If the null hypothesis of cross-group structural equality implied by the omnibus test could not be rejected, separate pairwise tests were not necessary because structural estimates were statistically equivalent across groups. However, if the hypothesis of the omnibus test was rejected, systematic pairwise tests were conducted to identify the source of structural cross-group noninvariance. The hypothesis that paths $p_{21}$ in each pair of groups *z* and *k* are invariant ($H_o$: $p_{21z} = p_{21k}$) is tested first, followed by the hypothesis that both paths, $p_{21}$ and $p_{32}$ are cross-group invariant ($H_o$: $p_{21z} = p_{21k}$; $p_{32z} = p_{32k}$). In each test, the chi-square difference ($\Delta\chi^2$) represents the difference between the structurally unconstrained model (denoted as ‘final scalar’) and the two constrained models. Table 16 reports the results of the omnibus tests and pairwise tests across white (W), Indian (I), Pakistani (P), Bangladeshi (B) and Black Caribbean (BC) groups. Evidence of moderation was assessed on the basis of the significance (p) of the chi-square difference test ($\Delta\chi^2$) for 1 (first test) and 2 (second test) degrees of freedom. In most but not all of the cases, it was also reflected in the change in CFI (ΔCFI) but not in the change of RMSEA (ΔRMSEA) due to the very well-fitting measurement models. Since bias in the multigroup chi-square was minimized by reducing sample discrepancy in the multigroup solution (see above), it was decided to base the decision regarding the rejection of the null hypothesis on the significance of $\Delta\chi^2$ at α (seventh column, table 16) adjusted by the Bonferroni correction. The last column of Table 15 reports this decision (R=reject or NR=not reject). The implications of rejection are discussed below.

Table 15: Cross-group comparison of the structural estimates of the model of pupils’ expectations between ages 14-16

| Hypothesis | **χ2** | df | p | **Δχ2** | df | **α** | CFI | **Δ**CFI | RMSEA | **Δ**RMSEA | Decision |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Final scalar[1] | 177.7 | 53 | 0.0 | - | - | - | 0.985 | - | 0.027 | - | |
| $p_{21k} = p_{21}$ | 217.3 | 57 | 0.0 | 39.4 | 4 | 0.00 | 0.981 | -0.014 | 0.030 | 0.003 | R |
| $p_{32k} = p_{32}$ | 219.8 | 61 | 0.0 | 41.9 | 8 | 0.00 | 0.981 | 0.000 | 0.029 | -0.001 | R |
| $p_{21W} \neq p_{21I}$ | 19.7 | 16 | ns | - | - | - | 0.999 | - | 0.011 | - | |
| $p_{21W} = p_{21I}$ | 23.3 | 17 | ns | 3.6 | 1 | 0.06 | 0.999 | 0.000 | 0.015 | 0.004 | NR |
| $p_{32W} = p_{32I}$ | 23.4 | 18 | ns | 3.7 | 2 | ns | 0.999 | 0.000 | 0.013 | -0.003 | NR |
| $p_{21W} \neq p_{21P}$ | 125.7 | 19 | 0.0 | - | - | - | 0.978 | - | 0.059 | - | |
| $p_{21W} = p_{21P}$ | 164.6 | 20 | 0.0 | 38.9 | 1 | 0.00 | 0.971 | -0.007 | 0.066 | -0.007 | R |
| $p_{32W} = p_{32P}$ | 164.8 | 21 | 0.0 | 39.1 | 2 | 0.00 | 0.971 | -0.007 | 0.066 | -0.007 | R |
| $p_{21W} \neq p_{21B}$ | 77.3 | 21 | 0.0 | - | - | - | 0.988 | - | 0.043 | - | |
| $p_{21W} = p_{21B}$ | 84.6 | 22 | 0.0 | 7.4 | 1 | 0.00 | 0.987 | -0.001 | 0.044 | 0.001 | R |
| $p_{32W} = p_{32B}$ | 85.6 | 23 | 0.0 | 8.4 | 2 | 0.00 | 0.987 | 0.000 | 0.043 | -0.001 | R |
| $p_{21W} \neq p_{21BC}$ | 67.2 | 21 | 0.0 | - | - | - | 0.989 | - | 0.041 | - | |
| $p_{21W} = p_{21BC}$ | 72.1 | 22 | 0.0 | 4.9 | 1 | 0.02 | 0.988 | 0.001 | 0.042 | 0.001 | R |
| $p_{32W} = p_{32BC}$ | 72.5 | 23 | 0.0 | 5.3 | 2 | 0.00 | 0.988 | 0.000 | 0.040 | -0.002 | R |
| $p_{21I} \neq p_{21P}$ | 30.5 | 16 | 0.0 | - | - | - | 0.995 | - | 0.026 | - | |
| $p_{21I} = p_{21P}$ | 43.9 | 17 | 0.0 | 13.4 | 1 | 0.00 | 0.991 | -0.004 | 0.037 | 0.011 | R |
| $p_{32I} = p_{32P}$ | 44.4 | 18 | 0.0 | 13.9 | 2 | 0.00 | 0.991 | 0.000 | 0.032 | -0.005 | R |
| $p_{21I} \neq p_{21B}$ | 13.8 | 18 | ns | - | - | - | 1.000 | - | 0.000 | - | |
| $p_{21I} = p_{21B}$ | 14.8 | 19 | ns | 1.0 | 1 | ns | 1.000 | 0.000 | 0.000 | 0.000 | NR |
| $p_{32I} = p_{32B}$ | 15.5 | 20 | ns | 1.7 | 2 | ns | 1.000 | 0.000 | 0.000 | 0.000 | NR |
| $p_{21I} \neq p_{21BC}$ | 27.9 | 18 | ns | - | - | - | 0.996 | - | 0.023 | - | |
| $p_{21I} = p_{21BC}$ | 28.3 | 19 | ns | 0.4 | 1 | ns | 0.996 | 0.000 | 0.021 | 0.001 | NR |
| $p_{32I} = p_{32BC}$ | 28.5 | 20 | ns | 0.6 | 2 | ns | 0.996 | 0.000 | 0.020 | 0.001 | NR |
| $p_{21P} \neq p_{21B}$ | 44.5 | 21 | 0.0 | - | - | - | 0.996 | - | 0.032 | - | |
| $p_{21P} = p_{21B}$ | 50.5 | 22 | 0.0 | 6.0 | 1 | 0.00 | 0.998 | 0.002 | 0.034 | 0.002 | R |
| $p_{32P} = p_{32B}$ | 52.8 | 23 | 0.0 | 8.3 | 2 | 0.00 | 0.988 | 0.000 | 0.034 | 0.000 | R |
| $p_{21P} \neq p_{21BC}$ | 59.1 | 21 | 0.0 | - | - | - | 0.981 | - | 0.043 | - | |
| $p_{21P} = p_{21BC}$ | 65.1 | 22 | 0.0 | 6.0 | 1 | 0.00 | 0.978 | -0.003 | 0.045 | 0.002 | R |
| $p_{32P} = p_{32BC}$ | 66.7 | 23 | 0.0 | 7.6 | 2 | 0.00 | 0.978 | -0.003 | 0.045 | 0.002 | R |
| $p_{21B} \neq p_{21BC}$ | 38.0 | 23 | 0.0 | - | - | - | 0.991 | - | 0.028 | - | |
| $p_{21B} = p_{21BC}$ | 38.1 | 24 | 0.0 | 0.1 | 1 | ns | 0.992 | 0.001 | 0.027 | -0.001 | NR |
| $p_{32B} = p_{32BC}$ | 38.1 | 25 | 0.0 | 0.1 | 2 | ns | 0.992 | 0.001 | 0.027 | -0.001 | NR |

[1]: ‘Final scalar’ refers to the final multigroup solution with metric and scalar invariance constraints in place, see last row, Table 14.
Note: **$\chi^2$** =chi-square; df=degrees of freedom; p=significance; **$\Delta\chi^2$**= chi-square difference; **α**=level of significance; CFI=comparative fit index (>0.95); RMSEA=Root mean square error or approximation (<0.05); **ΔCFI=change in CFI (≤ -0.01)**; **ΔRMSEA=change in RMSEA (≥ 0.01**6); R=reject $H_o$; NR=fail to reject the $H_o$ ;W=white; I=Indian; P=Pakistani, B=Bangladeshi and BC=Black Caribbean

Pupils’ expectations) showed very significant cross-group differences in both path $p_{21}$, representing change in expectations between ages 14 to 15 ($\Delta\chi^2 = 39.4$ (4) $p \leq 0.00$) and path $p_{32}$, representing change between ages 15 to 16 ($\Delta\chi^2 = 41.9$ (8) $p \leq 0.00$). The omnibus tests were therefore consistent with the hypothesis that change in expectations was moderated by maternal ethnicity. Some ethnic groups differed most in the change in expectations between ages

14 and 15 while others in the change between ages 15 to 16. The highest differences between ages 14 to 15 (path $p_{21}$) were found between the white pupils ($p_{21W}$=0.819) and their Pakistani ($p_{21P}$=0.527; $\Delta\chi^2 = 38.9$ (1), $p \leq 0.00$) and Bangladeshi ($p_{21I}$=0.655; $\Delta\chi^2 = 7.4$ (1), $p \leq 0.00$) peers. Indian young people ($p_{21I}$=0.690) also differed significantly from their Pakistani peers ($p_{21P}$=0.527; $\Delta\chi^2 = 13.4$ (1), $p \leq 0.00$). In turn, Pakistani pupils differed significantly from their Bangladeshi peers ($\Delta\chi^2 = 6.0$ (1), $p \leq 0.00$) during the same period.

Significant cross-group differences in expectations between age 15 and 16 (path $p_{32}$) centred mostly on differences between white pupils and their peers in all the other minority groups. Highly significant differences in both paths $p_{21}$ and $p_{32}$ were found between white pupils and their Pakistani ($\Delta\chi2 = 39.1$ (2), $p \leq 0.00$), Bangladeshi ($\Delta\chi2 = 8.4$ (2), $p \leq 0.00$), Black Caribbean pupils ($\Delta\chi^2 = 5.3$ (2), $p \leq 0.00$) but surprisingly, not their Indian ($\Delta\chi2 = 3.7$ (2), p = ns) peers. Pakistani pupils also differed significantly from their Bangladeshi ($\Delta\chi^2 = 8.3$ (2), $p \leq 0.00$) and their Black Caribbean ($\Delta\chi2 = 7.6$ (2), $p \leq 0.00$) peers. The above evidence suggests a complex picture of cross-group differences in the structural estimates of pupils' educational expectations between ages 14 to 16. White pupils differed most markedly from the rest of their peers both in having the lowest proportions of those planning to apply to university and be accepted if they applied and in being the group more likely to change their expectations from ages 14 to16. So, knowing the *level* of white pupils' expectations from ages 14 to 16 makes all the difference in understanding what their high temporal stability in expectations means. Significant differences in temporal stability were also found among the three South Asian groups as well as between them and their Black Caribbean peers. But these differences were smaller compared to those observed between the white and the rest of their minority peers. I will now place the above ethnic differences in temporal stability in proper context by analyzing differences in latent means and intercepts.

Latent *means* represent the *average* level of the latent construct in each group. Given metric and scalar cross-group invariance, they are error-free representations of between-group differences in a latent construct (Millsap, 2011). Latent *intercepts* in a repeated measures framework represent the contribution of the prior occasion on the next occasion. They represent the between-individual differences in the latent construct controlling for the effect of the prior occasion. A significant latent intercept suggests that the previous occasion has contributed to a significant *net* between-group latent difference in the next occasion. Thus, latent intercepts show to what extent change from a prior occasion to the next has resulted in significant between-individual differences in the latent construct of the next occasion. Thus, in an endogenous latent construct, its latent mean shows the level that has resulted in the latent construct due to the contribution of its prior occasion as shown by its latent intercept. Latent means and intercepts can be compared across ethnic groups. Put another way, these comparisons are actually tests of cross-group invariance in latent means and intercepts of a designated reference group and at least one comparison group. I describe the logic of such tests below.

Latent means and intercepts are unknown quantities of unobserved constructs. We cannot directly estimate the latent mean or the latent intercept of either the reference or the comparison groups. Sörbom (1974; 1978) has shown however, that we can estimate the *difference* in latent means and intercepts between the reference and the comparison groups if the measurement models of both groups are constrained to measurement invariance. Thus, latent means and intercepts represent scaled point *differences* between the latent mean(s) and latent intercept(s) of the reference group and those of the comparison groups. They test the hypothesis of cross-group equality in factor means ($H_o$: $\boldsymbol{\mu}_{\kappa} = \boldsymbol{\mu}$) and factor intercepts ($H_o$: $\boldsymbol{\kappa}_{\kappa} = \boldsymbol{\kappa}$). A difference in latent factor means and intercepts is statistically significant if the ratio of the produced difference to its standard error exceeds the critical ratio (CR) of 1.98 at $p = 0.05$. A Bonferroni correction was implemented to adjust for Type I error, and as a result, the p level was reduced to $p = 0.0125$, as described above.

With 'strong' measurement invariance, between-group differences stem only from the latent construct means. Significant cross-group differences in latent means and intercepts therefore represent error-free evidence of moderation by the group differentiating factor, i.e., maternal ethnicity. This permitted comparison of latent means and intercepts both longitudinally and across groups, although in some past research, latent means have been compared solely on the basis of metric or 'weak' factorial invariance (Schaie *et al.*, 1998).

Comparison of latent means and intercepts also involved $c = k^*(k-1)/2$ comparisons where $k$ = groups in the analysis. A Bonferroni correction was also implemented in this case. In the literature, such comparisons commence with full metric and scalar measurement invariance imposed on both measurement parts of the models in the pairwise comparison. However, this level of invariance must be supported by the data, as was the case in the present analysis, not simply imposed on the measurement model (Millsap, 2013, personal communication). Sörbom (1974; 1978) suggested *theta* ($\theta$) invariance should be imposed as well (invariant uniquenesses). However, this requirement for 'strict' factorial invariance (Meredith, 1993) was not implemented in this analysis because 'strict' factorial invariance is not a prerequisite to proceed with the comparison of latent means and intercepts (Millsap, 2011; Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000).

In order to proceed with the pairwise comparisons of latent means, each autoregressive SEM was respecified as a CFA autoregressive longitudinal model, replacing the longitudinal paths $p_{21}$ and $p_{32}$ with factor covariances ($\mathbf{\Phi}_{21}$, $\mathbf{\Phi}_{32}$). In this way, the software estimates differences in factor means of all latent constructs but not latent intercepts since there is no $\mathbf{\Gamma}$ matrix (factor covariances between exogenous and endogenous factors). Following standard procedure, the latent means of the reference group were constrained to zero while those of the comparison group were freely estimated. To estimate latent intercepts, the CFA model was respecified as an autoregressive longitudinal SEM, replacing the factor covariances with the longitudinal paths $p_{21}$ and $p_{32}$. The software now estimated differences in factor means of only the exogenous factors (which were identical to those obtained for the same factor in the prior estimation step) plus differences in factor intercepts for every endogenous factor, since in the case of SEM, both the $\mathbf{B}$ and the $\mathbf{\Gamma}$ matrices are estimated. The level of measurement invariance remained the same but the latent intercepts of the reference model were constrained to zero while those of the comparison group were freely estimated. Table 16 shows the results from the hypothesis tests of equality of latent means and intercepts and their standard errors for the model of pupils' educational expectations. The reference group in each comparison is noted in bold. The white group is first compared to all others followed by Indian group which is compared next to the remaining three ethnic groups. This is followed by the Pakistani group which is compared next to the remaining 2 groups and last by the Bangladeshi which compared only to the remaining Black Caribbean group, resulting in $c = k^*(k-1)/2 = 5^*(5-1)/2=10$ comparisons. Differences in latent means of the exogenous factor ($\boldsymbol{\mu}_1$) are reported first, followed by the differences in factor intercepts of the first endogenous factor ($\boldsymbol{\kappa}_2$), whose differences in factor means ($\boldsymbol{\mu}_2$) are reported next, followed by the differences in factor intercepts ($\boldsymbol{\kappa}_3$) and factor means ($\boldsymbol{\mu}_3$) of the second endogenous factors.

Table 16: Comparison of factor means and intercepts in the model of pupils' educational expectations across all ethnic minority groups.

| | YPEX1 latent means $\mu_1$ | | | YPEX2 latent intercepts $\kappa_2$ | | | YPEX2 latent means $\mu_2$ | | | YPEX3 latent intercepts $\kappa_3$ | | | YPEX3 latent means $\mu_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ethnic group | $\mu_1$ | SE | p | $\kappa_2$ | SE | p | $\mu_2$ | SE | p | $\kappa_3$ | SE | p | $\mu_3$ | SE | p |
| **White (n=1000)** | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | |
| Indian (n=751) | 0.499 | 0.036 | 0.000 | 0.080 | 0.023 | 0.00 | 0.506 | 0.033 | 0.00 | 0.170 | 0.022 | 0.00 | 0.602 | 0.034 | 0.00 |
| Pakistani (n=642) | 0.345 | 0.038 | 0.000 | 0.045 | 0.028 | 0.250 | 0.335 | 0.036 | 0.00 | 0.065 | 0.025 | 0.01 | 0.343 | 0.038 | 0.00 |
| Bangladeshi (n=484) | 0.234 | 0.042 | 0.000 | 0.067 | 0.031 | 0.029 | 0.261 | 0.040 | 0.00 | 0.090 | 0.028 | 0.00 | 0.312 | 0.040 | 0.00 |
| BCaribbean (n=324) | 0.263 | 0.050 | 0.000 | -0.064 | 0.039 | 0.300 | 0.160 | 0.047 | 0.00 | 0.139 | 0.037 | 0.00 | 0.281 | 0.050 | 0.00 |
| **Indian (n=751)** | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | |
| Pakistani (n=642) | -0.192 | 0.035 | 0.000 | -0.056 | 0.028 | 0.04 | -0.209 | 0.033 | 0.00 | -0.035 | 0.025 | 0.232 | -0.199 | 0.036 | 0.00 |
| Bangladeshi (n=484) | -0.304 | 0.039 | 0.000 | -0.041 | 0.031 | 0.350 | -0.281 | 0.038 | 0.00 | -0.015 | 0.028 | 0.450 | -0.231 | 0.038 | 0.00 |
| BCaribbean (n=324) | -0.276 | 0.047 | 0.000 | -0161 | 0.039 | 0.000 | -0.384 | 0.045 | 0.00 | 0.030 | 0.037 | 0.321 | -0.262 | 0.048 | 0.00 |
| **Pakistani (n=642)** | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | |
| Bangladeshi (n=484) | -0.111 | 0.040 | 0.004 | 0.003 | 0.031 | 0.892 | -0.078 | 0.039 | 0.045 | 0.027 | 0.028 | 0.257 | -0.029 | 0.040 | 0.471 |
| BCaribbean (n=324) | -0.086 | 0.018 | 0.070 | -0.120 | 0.038 | 0.002 | -0.184 | 0.046 | 0.000 | 0.072 | 0.036 | 0.05 | 0.00 | -0.060 | 0.220 |
| **Bangladeshi (n=484)** | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | | 0.00 | | |
| BCaribbean (n=324) | -0.008 | 0.051 | 0.818 | -0.099 | 0.039 | 0.010 | -0.111 | 0.048 | 0.021 | 0.094 | 0.037 | 0.01 | 0.012 | 0.051 | 0.809 |

Note: YPEX-3; occasions 1-3 in the model of pupils' educational expectations (see Figure 1); p= significance; SE=standard error. Reference groups are noted in bold; n=sample size

Table 16 suggests that at age 14, all groups held significant positive differences in their latent means in educational expectations relative to their white peers. The highest differences were noted in the Indian ($\Delta_{\mu I} = 0.499$, $p \leq 0.00$) pupils, followed by their Pakistani ($\Delta_{\mu P} = 0.345$, $p \leq 0.00$), Black Caribbean ($\Delta_{\mu BC} = 0.263$, $p \leq 0.00$) and Bangladeshi ($\Delta_{\mu B} = 0.234$, $p \leq 0.00$) peers. All of these gaps widened at age 15 and even more so at age 16 relative to white pupils. However, the gaps in latent mean differences of minority pupils relative to their white peers widened very differently in each minority group. Based on the latent intercepts for each group at ages 15 and 16, Indian pupils had the highest net gains in expectations thus maintaining the biggest gaps in latent mean differences in expectations relative to all other groups between ages 14 to 16. Differences in their expectations increased most dramatically relative to all other groups between ages 15 to 16. Pakistani pupils also held consistent positive gaps in latent mean differences in expectations relative to their white peers. But contrary to their Indian peers, these gaps hardly changed between ages 14 to 16 (see Table 16, columns, $\mu_1$, $\mu_2$ and $\mu_3$ for Indian and Pakistani pupils). Bangladeshi pupils widened their positive gaps in latent mean differences in expectations between ages 14 to 15 and still more between ages 15 to 16. Like their Indian peers, they increased their latent mean differences more during ages 15 to 16. However, comparing the latent intercepts in the Indian and Bangladeshi groups at ages 15 and 16, Indian pupils widened their positive gaps in expectations relative to their white peers much *faster* than did Bangladeshi pupils. This is easily confirmed by the nonsignificant negative latent intercepts for Bangledeshi pupils at age 15 ($\Delta_{\kappa B} = -0.041$, p = ns) and 16 ($\Delta_{\kappa B} = -0.015$, p = ns).

Black Caribbean pupils were remarkable in being the only group of pupils to start off at age 14 with a positive gap in their latent mean expectations relative to their white peers ($\Delta_{\mu BC} = 0.260$, p ≤ 0.00); to lower their expectations, narrowing this gap considerably at age 15 ($\Delta_{\mu BC}$ = 0.160, p ≤ 0.00); and more than regain that advantage relative to their white peers by increasing their latent mean differences again at age 16 ($\Delta_{\mu BC}$ = 0.281, p ≤ 0.00). This peculiar curve was confirmed by their nonsignificant negative latent intercept at age 15 ($\Delta_{\kappa BC}$ = -0.064, p = ns) and their significant positive latent intercept difference at age 16 ($\Delta_{\kappa BC}$ = 0.134, p ≤ 0.00). In terms of the rate of increase in their latent mean expectations, Black Caribbean pupils caught up with their Pakistani and Bangladeshi peers at age 16. This was also confirmed by the Black Caribbean latent intercept differences when those two groups were the reference groups (see Table 16, columns $\kappa_2$ and $\kappa_3$, last six rows). The above evidence is entirely consistent with the observed cross-group noninvariance of the paths $p_{21}$ and $p_{32}$ in the model of educational expectations discussed above.

However we are in a position now to interpret this cross-group noninvariance in terms of the underlying *levels* of expectations. Cross-group structural non-invariance existed particularly in paths $p_{21}$ and $p_{32}$ because Indian pupils increased their expectations faster than any other group between ages 14 to 16. Pakistani pupils were different from their Indian peers because Pakistani pupils' expectations remained much more stable (but considerably lower) than those of their Indian peers. Bangladeshi pupils were different because although their expectations increased over time, the gaps between them and their other South-Asian peers varied. Finally Black Caribbean pupils were different in that no other group showed a slump in expectations from ages 14 to 15 and a fast recovery from ages 15 to 16 relative to their white peers.

I complete the analysis of longitudinal change in pupils' educational expectations by examining whether the above cross-group structural non-invariance in paths $p_{21}$ and $p_{32}$ also means that the model of pupils' expectations across those ethnic groups is in equilibrium. Equilibrium refers to a condition when the causal system exhibits temporal stability of patterns of covariance and variance (Dwyer, 1983). It is in fact a test of longitudinal factor variance invariance and must be performed in addition to those of stationarity, correlated errors and measurement invariance (Cole and Maxwell, 2003), which were already addressed in this analysis. It has been suggested that a causal system may be stationary but not in equilibrium. Alternatively, it may be in equilibrium but not stationary (Cole and Maxwell, 2003). Thus, longitudinal or cross-group equality in the structural parameters (here, dependence paths $p_{21}$ and $p_{32}$) does *not* necessarily imply equilibrium. Table 17 presents the results of the tests for longitudinal factorial variance equivalence (indicating presence or absence of equilibrium) in each ethnic group.

Table 17: Tests of equilibrium for the model of pupils' expectations for each ethnic minority group

| | Factor variance at each occasion | | | Hypothesis | $\chi^2$ | df | p | $\Delta\chi^2$ | df | p | Decision |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ethnic groups | YPEX1 | YPEX2 | YPEX3 | | | | | | | | |
| White (n=1000) | .784 | .895 | .945 | v1 ≠ v2 ≠ v3 | 15.1 | 9 | ns | - | - | - | - |
| | | | | v1 = v2 = v3 | 39.1 | 11 | .00 | 24 | 2 | .00 | R |
| Indian (n=751) | .290 | .381 | .384 | v1 ≠ v2 ≠ v3 | 2.3 | 7 | ns | - | - | - | - |
| | | | | v1 = v2 = v3 | 15.0 | 9 | ns | 12.7 | 2 | .00 | R |
| Pakistani (n=642) | .345 | .363 | .429 | v1 ≠ v2 ≠ v3 | 14.5 | 8 | ns | - | - | - | - |
| | | | | v1 = v2 = v3 | 21.3 | 10 | .02 | 6.8 | 2 | .03 | R |
| Bangladeshi (n=484) | .369 | .432 | .395 | v1 ≠ v2 ≠ v3 | 11.2 | 10 | ns | - | - | - | - |
| | | | | v1 = v2 = v3 | 14.1 | 12 | ns | 2.1 | 2 | ns | NR |
| Black Caribbean (n=324) | .409 | .377 | .449 | v1 ≠ v2 ≠ v3 | 18.6 | 9 | .02 | - | - | - | - |
| | | | | v1 = v2 = v3 | 19.0 | 11 | ns | .04 | 2 | ns | NR |

Note:YPEX1-3=pupils' expectations at age 14, 15 and 16; $\chi^2$ =chi-square; df=degrees of freedom; p=significance; $\Delta\chi^2$= chi-square difference; CFI=comparative fit index (>0.95); RMSEA=Root mean square error or approximation (<0.05); R=reject $H_o$; NR=fail to reject the $H_o$

Bringing together the evidence presented by Tables 13 and 17 suggests that the causal system of pupils' educational expectations was neither in equilibrium nor stationary for the white and Pakistani pupils. It was stationary but not in equilibrium for their Indian peers, while it was both stationary and in equilibrium for their Bangladeshi and Black Caribbean peers. There are quite important substantive implications regarding pupils' expectations based on the above analysis. Maternal ethnicity moderated differently not only the change parameters in expectations from ages 14 to 15 and between ages 15 to 16 but also the change in the latent mean levels of pupils' expectations. In other words, maternal ethnicity appears to moderate the rate of change in expectations from ages 14 to 15 and from ages 15 to 16 (assessed by significant differences in the dependence paths across groups) but also the extent to which each prior occasion of pupils' educational expectations impacts on the next (assessed by the significant differences in latent intercepts across groups). In addition, maternal ethnicity moderated the latent means in expectations (assessed by the significant differences in latent means across groups). Finally, maternal ethnicity seems to moderate the extent to which the causal system of pupils' educational expectations between ages 14 to 16 has reached equilibrium or alternately, is still in the process of development. The evidence suggests that expectations for the white, Indian and Pakistani pupils are still developing, while they have reached equilibrium for their Bangladeshi and Black Caribbean peers. This conclusion might seem at odds with the fact that the causal system in the Indian pupils was stationary, suggesting that the rate of change between occasions remained similar. But the similar dependence between occasions simply suggests that expectations at age 15 are influenced by expectations at age 14 as much as expectations at age 16 are influenced by expectations at age 15. The non-invariant variances in the latent construct across ages 14, 15 and 16 however mean that the attitudes towards applying to university and of being accepted if applying have not reached stability across time. For the white pupils this seems to occur between ages 14-16, while for the Indian pupils mostly between ages 14 to 15 and the Pakistani pupils between ages 15 and 16. This moderation creates and maintains significant gaps in the latent mean levels of pupils' expectations that varied in each group over time.

## 7. Discussion and conclusions

This paper addressed two questions. First whether pupils' educational expectations changed significantly over the time window of ages 14 to 16. Second, whether this change was moderated by maternal ethnicity. The analysis demonstrated that strong longitudinal and cross-group invariance was supported by the data, permitting comparisons of structural estimates both longitudinally and across ethnic groups. This analysis contributes to past UK research on pupils' educational expectations because it subjected the measurement assumptions required for such comparisons to rigorous tests. These assumptions were (a) stationarity of longitudinal structural parameters, (b) equilibrium and (c) longitudinal and cross-group invariance (d) extraction of method and trait variance in the correlated indicator error structure. Because the psychometric framework permitted direct test of these assumptions, observed change in the structural estimates of the model of pupil's educational expectations was true change. This analysis also controlled against bogus moderator effects by reducing the likelihood of biased group chi-square resulting from very discrepant sample sizes as well as applying a Bonferonni correction of the significance level of the chi-square difference tests.

Thus, the analysis showed that there were significant cross-group differences in the change of young people's educational expectations over time. Expectations increased over time most dramatically in the Indian pupils relative to all other groups, while they remained consistently the lowest among white pupils. This does not mean of course that there are no white pupils that excel in performance and that they maintain high expectations to apply and be accepted if they apply to university. However, on the *average* level of the latent dimension of such academic expectations, white pupils are found to develop their expectations much slower than their South Asian and Black Caribbean peers do. *Level* of expectations is critical in this connection. Indian pupils resemble white pupils in that in both groups, expectations between ages 15 to 16 have the identical dependence path (see Table 11). But for Indian pupils, this means that their much higher expectations at age 15 strongly determine their expectations at age 16. The same dependence between white pupils' generally much lower expectations at age 15 explains why white pupils' expectations remain the lowest at age 16 as well.

A related significant finding was that earlier expectations determined later expectations for all groups. Therefore, in any analysis aiming to explain expectations associated with a particular age, either performed under a psychometric framework or not, it is important not to ignore the influence of expectations prior to those expectations we are trying to explain. Expectations measured only at a single point in time will, in other words, offer misleading estimates if the design fails to control for prior expectations. In the literature, this is described as a great potential confounder, and ignoring it will severely bias all estimates associated with the target expectations (Cole and Maxwell, 2003; Gollob and Reichardt, 1985).

It was also shown that earlier expectations impact on later expectations both directly and indirectly. The present analysis in fact showed that while the direct effect of early expectations on later expectations wanes with time, and the strongest direct influence on expectations at age 16 are expectations at age 15, the indirect effect of expectations at age 14 on expectations at age 16 remains very strong and significant. In practical terms this means that information related to a pupils' decision to apply to university or hope that he/she will get in if apply obtained at age 15 will force expectations to change more after that point in time. Fumagalli's (2012) suggestion that pupils' expectations tended to change immediately after GCSE results were known at age 15 seems therefore to gain some support. However, the present analysis also showed that information gained much earlier at age 14 also exerts important *indirect* effects on expectations at age 16. Although such a conclusion seems plausible, this analysis has precisely measured this indirect longitudinal influence. It showed that although this indirect influence can never be as strong as the direct influence of the immediately prior expectations, it nevertheless remains a very significant influence on age-16 expectations. In that respect, it is probable that Key Stage 3 exams sat at age 14, and most likely Key Stage 2 exams sat at age 11, (see.Anders and Micklewright, 2013) also contribute to the faster change in expectations between ages 15 to 16, across all minority ethnic groups. If academic performance is moderated by maternal ethnicity, which is a fact the relevant literature strongly implies or clearly suggests (Abbas, 2002; Anders and Micklewright, 2013; Modood, 2004; Modood *et al.*, 1997; Rothon, 2005; Rumbaut, 1994; Siraj-Blatchford, 2010; Strand, 2007)], then we might find it easier to explain what causes expectations to change as they do between ages 14 to 16. The present research however contributed to past research on pupils' expectations in offering an analysis of differences in the longitudinal change in expectations of minority ethnic pupils between ages 14 to 16 that had never before been systematically undertaken.

The major finding of the analysis in that respect is that maternal ethnicity moderates change in expectations quite significantly. This moderation is quite complex. It impacts on expectations in at least four ways: (a) by affecting the rate of change between ages 14 to 15 and 15 to 16; (b) by affecting the extent to which prior expectations exert a direct net impact on later expectations;

(c) by affecting the strength of indirect influences of prior expectations on later expectations; and (d) by affecting the mean levels of the latent dimension capturing pupils’ expectations related to university study. In this respect, white and Pakistani pupils’ average expectations were neither stationary nor in equilibrium. But the direction in which these expectations changed was different. Pakistani pupils’ expectations were significantly higher than those of their white peers at age 14 and remained so at age 15 and 16 with higher rates of change in that period. So, the relative flux and development in their expectations points to a steady improvement. This is not suggested at all by their white peers. In their case, and given that the gaps in average expectations between them and their South Asian and Black Caribbean peers significantly increased with time, development in expectations means further and faster divergence from the expectations of their peers. Indian, Bangladeshi and Black Caribbean pupils’ expectations were much more stable over time than those of their white and Pakistani peers. But stability means different things depending on whether expectations were on average high or low. Indian pupils, maintained consistently the highest expectations between ages 14 to 16, so in their case, their stability suggests that high expectations at age 14 will be associated with high expectations at age 15 and 16. White pupils, on the other hand, have the lowest expectations relative to all their other peers at age 14 and gaps between white pupils and all the rest tended to increase at age 16. In their case therefore, change in expectations leads to still bigger gaps in expectations in favour of South Asian and Black Caribbean groups.

Expectations changed over time affecting the gaps in expectations differentially among ethnic minority groups. While maternal ethnicity moderated development of expectations across ethnic minorities in England, earlier expectations at age 14 remained the most important influence on later expectations at age 16. This points to the substantive importance of pupils’ expectations at age 14 regarding university study, three or four years earlier than the actual application will take place. This suggests that the decision to attend university is the result of a complex longitudinal process involving the home and school and future research must uncover these routes of influence.

## Bibliography


Abbas, T. (2002). 'The home and the school in the educational achievements of South Asians'. *Race, Ethnicity and Education,* 5 (3), 292-316.

Aguinis, H. (1995). 'Statistical power problems with moderated multiple regression in management research'. *Journal of Management Research,* 21, 1141-1158.

Aguinis, H., Petersen, S. A. and Pierce, C. A. (1999). 'Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables'. *Organizational Research Methods,* 2, 315-339.

Alexander, K. and Eckland, B. K. (1975). 'Contextual Effects in the High School Attainment Process'. *American Sociological Review* 40 (3), 402-416.

Anders, J. and Micklewright, J. (2013). *Teenagers' expectations of applying to university: how do they change?* London: Institute of Education.

Arbuckle, J. L. (2009). *AMOS 18.0 User's Guide*. Chicago IL: Amos Development Corporation.

Arbuckle, J. L. (2011). *IBM-SPSS Amos20 User’s Guide*: IBM Corporation.

Babakus, E., Ferguson, C. E. J. and Jöreskog, K. G. (1987). 'The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions.'. *Journal of Marketing Research,* 24, 222-228.

Bagozzi, R. P. and Yi, Y. (1988). 'On the evaluation of structural equation models'. *Journal of the academy of marketing science,* 16 (1), 74-94.

Baron, R. M. and Kenny, D. A. (1986). 'The mediator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations'. *Journal of Personality and Social Psychology,* 51 (6), 1173 - 1182.

Basit, T. N. (2012). '‘My parents have stressed that since I was a kid’: Young minority ethnic British citizens and the phenomenon of aspirational capital'. *Education, Citizenship and Social Justice* 7(2), 129-143.

Bentler, P. M. (1990). 'Comparative fit indexes in structural models'. *Psychological Bulletin* 107, 238-246.

Bentler, P. M. and Bonett, D. G. (1980). 'Significance tests and goodness of fit in the analysis o covariance structures'. *Psychological Bulletin,* 88, 588-606.

Bentler, P. M. and Chou, C.-P. (1987). 'Practical issues in structural modeling'. *Sociological Methods and Research,* 16, 78-117.

Bhandari, M., Whang, W., Kuo, J. C., Devereaux, P. J., Sprague, S. and Tornetta Iii, P. (2003). 'The risk of false-positive results in orthopaedic surgical trials'. *Clinical orthopaedics and related research,* 413, 63-69.

Bijleveld, C. C. J. H., van der Kamp, L. J. T., Mooijaart, A., van der Kloot, W. A., van der Leeden, R. and van der Burg, E. (1998). *Longitudinal data analysis: Designs, models and methods*. London Sage Publications.

Bland, J. M. and Altman, D. G. (1995). 'Multiple significance tests: the Bonferroni method'. *BMJ: British Medical Journal,* 310, 170.

Blunch, N. J. (2010). *Introduction to structural equation modelling using SPSS and AMOS*. London Sage Publications, LTD.

Bollen, K. A. (1986). 'Sample size and Bentler and Bonett's nonnormed fit index'. *Psychometrika,* 51 (3), 375-377.

Bollen, K. A. (1989a). *Structural equations with latent variables*. New York John Wiley and Sons.

Bollen, K. A. (1989b). 'A new incremental fit index for general equation models'. *Sociological Methods and Research,* 17, 303-316.

Bollen, K. A. (1990). 'Overall fit in covariance structure models: two types of sample size effects'. *Psychological Bulletin,* 107 (2), 256-259.

Bollen, K. A. and Long, J. S. (1993). *Testing structural equation models*. London: Sage.

Bond, R. and Saunders, P. (1999). 'Routes of success: influences on the occupational attainment of young British males'. *British Journal of Sociology,* 50 (2), 217-249.

Bourdieu, P. and Passeron, J. C. (1977). *Reproduction in Education, Society and Culture*. Beverly Hills: Sage.

Bowers, E. P., Li, Y., Kiely, M. K., Brittian, A., Lerner, J. V. and Lerner, R. M. (2010). 'The five Cs model of positive youth development: A longitudinal analysis of confirmatory factor structure and measurement invariance'. *Journal of Youth and Adolescence,* 39 (7), 720-735.

Breen, R. and Goldthorpe, J. H. (1997). 'Class, Mobility and Merit: the experience of two British birth cohorts'. *European Sociological Review,* 17 (2), 81-101.

Bronfenbrenner, U. (2005). *Making Human Beings Human: Bioecological Perspectives on Human Development*. Thousand Oaks CA: Sage.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: The Guildford Press.
Browne, M. W. and Cudeck, R. (1993). 'Alternative ways of assessing model fit'. In B. K.A. and L. S.J. (Eds), *Testing structural equation models*. Newbury Park, CA Sage Publications Inc.
Buchmann, C. and Dalton, B. (2002). 'Interpersonal Influences and Educational Aspirations in 12 Countries: The Importance of
Institutional Context'. *Sociology of Education,* 75 (2), 99-122.
Byrne, B. M. (2004). 'Testing for multigroup invariance using AMOS graphics: A road less traveled'. *Structural Equation Modeling,* 11 (2), 272-300.
Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming, 2nd ed*. New York, NY Routledge Taylor and Francis Group.
Byrne, B. M., Shavelson, R. J. and Muthén, B. (1989). 'Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance'. *Psychological Bulletin,* 105 (3), 456-466.
Byrne, B. M. and Stewart, S. M. (2006). 'Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process'. *Structural Equation Modeling,* 13 (2), 287-321.
Chan, D. (1998). 'The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM)'. *Organizational Research Methods,* 1 (4), 421-483.
Chen, F. F. (2007). 'Sensitivity of goodness of fit indexes to lack of measurement invariance'. *Structural Equation Modeling,* 14 (3), 464-504.
Chen, F. F. (2008). 'What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research'. *Journal of personality and social psychology,* 95 (5), 1005-1018.
Cheung, G. W. and Rensvold, R. B. (1999). 'Testing factorial invariance across groups: A reconceptualization and proposed new method'. *Journal of Management,* 25 (1), 1-27.
Cheung, G. W. and Rensvold, R. B. (2002). 'Evaluating goodness-of-fit indexes for testing measurement invariance'. *Structural equation modeling,* 9 (2), 233-255.
Cole, D. A. and Maxwell, S. E. (2003). 'Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling'. *Journal of Abnormal Psychology,* 112 (4), 558-577.
Coleman, J. S. (1988). 'Social Capital and the Creation of Human Capital'. *American Journal of Sociology,* 94, 95-120.
Croll, P. and Attwood, G. (2013). 'Participation in higher education: aspirations, attainment and social background'. *British Journal of Educational Studies,* 61 (2), 187-202.
DfES. (2006). Making good progress: how can we help every pupil to make good progress at school? (Vol. 2). Nottingham: Department of Education and Skills
Dimitrov, D. M. (2010). 'Testing for factorial invariance in the context of construct validation'. *Measurement and Evaluation in Counseling and Development,* 43 (2), 121-149.
Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. New York: Oxford University Press
Edwards, J. R. and Lambert-Schurer, L. (2007). 'Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis'. *Psychological Methods,* 12 (1), 1-22.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. and Strahan, E. J. (1999). 'Evaluating the use of exploratory factor analysis in psychological research'. *Psychological methods,* 4 (3), 272.
Feinstein, L. (2003). 'Inequality in the early cognitive development of British children in the 1970 cohort'. *Economica,* 70 (277), 73-97.
Feinstein, L., Duckworth, K. and Sabates, R. (2004). *A Model of the inter-generational transmission of educational success*. London: Centre for Research on the Wider Benefits of Learning, Report No 10.
Finch, J. F., West, S. G. and MacKinnon, D. P. (1997). 'Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models'. *Structural Equation Modeling: A Multidisciplinary Journal*, 4 (2), 87-107.
Frazier, P. A., Tix, A. P. and Barron, K. E. (2004). 'Testing moderator and mediator effects in counseling psychology'. *Journal of Counceling Psychology*, 51 (1), 115-134.
Fumagalli, E. (2012). 'Great expectations. Channels and barriers to university education'. [http://users.unimi.it/brucchiluchino/conferences/2012/fumagalli.pdf](http://users.unimi.it/brucchiluchino/conferences/2012/fumagalli.pdf).
Furlong, A. (1986). 'Schools and the structure of female occupational aspirations'. *British Journal of Sociology of Education*, 7 (4), 367-377.
Furlong, A. and Biggard, A. (1999). 'Framing 'choices': a longitudinal study of occupational aspirations among 13- to 16-year olds'. *Journal of Education and Work,* 12 (21-35).
Goldthorpe, J. H. (2007). '"Cultural Capital": Some Critical Observations'. *Sociologica,* 2.
Goldthorpe, J. H. and Breen, R. (2000). 'Explaining Educational Differentials: Towards a Formal Rational Action Theory '. In J. H. Goldthorpe (Ed.), *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.
Gollob, H. F. and Reichardt, C. S. (1985). *Building time lags into causal models of cross-sectional data*. Paper presented at the Proceedings of the Social Statistics Section of the American Statistical Association.
Gollob, H. F. and Reichardt, C. S. (1987). 'Taking account of time lags in causal models'. *Child Development,* 58 (1), 80-92.
Gottfredson, L. S. (1981). 'Circumscription and Compromise: A Developmental Theory of Occupational Aspirations'. *Journal of Counseling Psychology Monograph*, 28 (6).
Gottfredson, L. S. (1996). 'Gottfredson's theory of circumscription and compromise'. In B. D. and L. Brooks (Eds), *Career Choice and Development*. San Francisco, CA: Jossey-Bass.
Goyette, K. A. (2008). 'College for some to college for all: social background, occupational expectations, and educational expectations over time'. *Social Sicience Research* 37, 461-484.
Gu, F., Preacher, K. J. and Ferrer, E. (2014). 'A state space modeling approach to mediation analysis'. *Journal of Educational and Behavioral statistics,* 39 (2), 117-143.
Gulliksen, H. and Tukey, J. W. (1958). 'Reliability for the law of comparative judgment'. *Psychometrika* 23, 95-110.
Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation and conditional process modeling [white paper], *An introduction to mediation, moderation and conditional process analysis: a regression approach*. New York: Guildford Press.
Hayes, A. F. (2013). *Introduction to mediation, moderation and conditional process analysis: a regression-based approach*. New York: The Guilford Press.

Hertzog, C. and Schaie, K. W. (1988). 'Stability and change in adult intelligence: II. Simultaneous analysis of longitudinal means and covariance structures'. *Psychology and Aging,* 3 (2), 122-130.
Hofferth, S. L. (1980). *High School, Occupational Choice, and Sex Equity*. Washington, DC: Urban Institute, National Institute of Education (DHEW).
Holmbeck, G. N. (1997). 'Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: examples from the child-clinical and pediatric psychology literatures'. *Journal of Counceling and Clinical Psychology,* 65, 599-610.
Hu, L. T. and Bentler, P. M. (1993). *Fit indices in covariance structural equation modeling*. New York: Free Press.
Hu, L. T. and Bentler, P. M. (1998). 'Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification'. *Psychological methods,* 3 (4), 424-453.
Hu, L. T. and Bentler, P. M. (1999). 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives'. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1-55.
IBM. (2011). 'Statistical Package for the Social Sciences (SPSS) version 20'.
Jaccard, J. and Turrisi, R. (2003). *Interaction effects in multiple regression*. (2nd ed.). Thousand Oaks, CA: Sage Publications.
Jaccard, J. and Wan, C. K. (1995). 'Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: multiple indicator and structural eqaution approaches'. *Psychological Bulletin,* 117, 433-465.
Jacob, B. A. and Wilder, T. (2010). *Educational expectations and attainment* (15683). Cambridge, MA: National Bureau of Economic Research.
Jerrim, J. (2011). *Disadvantaged children's 'low' eductional expectations: are the US and UK really so different to other industrialized nations?* Department of Quantitative Social Science: Institute of Education, University of London.
Jöreskog, K. G. (1969). 'A general approach to confirmatory maximum likelihood factor analysis'. *Psychometrika* 34, 183-202.
Jöreskog, K. G. (1971). 'Statistical analysis of sets of congeneric tests'. *Psychometrika,* 36 (2), 109-133.
Jöreskog, K. G. (1974). 'Analyzing psychological data by structural analysis of covariance matrices', *Contemporary Developments in Mathematical Psychology - Volume II*. San Fransisco: W.H. Freeman Co.
Jöreskog, K. G. (1993). 'Testing structural equation models'. In K. A. Bollen and S. J. Long (Eds), *Testing structural equation models* Newbury Park, CA: Sage Publications Inc.
Jöreskog, K. G. and Sörbom, D. (1984). *Lisrel-VI user's guide (3rd ed)*. Mooresville, IN Scientific Software.
Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: User's reference guide*. Los Angeles, CA: Scientific Software International.
Kao, G. and Thompson, J. S. (2003). 'Racial and ethnic stratification in educational achievement and attainment'. *Annual Review of Sociology,* 29, 417-442.
Kao, G. and Tienda, M. (1998). 'Educational Aspirations of Minority Youth'. *American Journal of Education,* 106 (3), 349-384.
Kaplan, D. and George, R. (1995). 'A study of the power associated with testing factor mean differences under violations of factorial invariance'. *Structural equation modeling,* 2 (2), 101-118.
Kelly, A. J. D. (1989). *Ethnic Identification, Association and Redefinition - Muslim Pakistanis and Greek Cypriots In Britain*

Kerckhoff, A. C. (1977). 'The Realism of Educational Ambitions in England and the United States'. *American Sociological Review,* 42 (4), 563-571.
Kline, R. B. (2005). *Principles and practice of structural equation modeling (2nd ed)*. New York The Guildford Press.
Kraemer, H. C., Wilson, G. T., Fairburn, C. G. and Agras, W. S. (2002). 'Mediators and moderators of treatment effects in randomized clinical trials'. *Archives of General Psychiatry*, 59 (10), 877.
Krauss, I. (1964). 'Sources of Educational Aspirations Among Working-Class Youth'. *American Sociological Review,* 29 (6), 867-879.
Leibnowitz, A. (1974). 'Home investments in children'. *The Journal of Political Economy,* 82 (2 Part 2: Marriage, Family Human Capital, and Fertility (Mar. - Apr., 1974), S111-S131.
Little, T. D. (1997). 'Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues'. *Multivariate Behavioral Research*, 32 (1), 53-76.
Little, T. D., Preacher, K. J., Selig, J. P. and Card, N. A. (2007). 'new developments in latent variable panel analyses of longitudinal data'. *International Journal of behavioral development*, 31 (4), 357-365.
MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York Taylor and Francis.
Marjoribanks, K. (1999). 'Environmental and individual influences on adolescents' aspirations: a moderation-mediation model.'. *Learning Environments Research*, 2, 43-64.
Marjoribanks, K. (2003a). 'Family Background, Individual and Environmental Influences, Aspirations and Young Adults' Educational Attainment: a follow-up study'. *Educational Studies*, 29 (2), 233-242.
Marjoribanks, K. (2003b). 'Learning Envorinments, Family Contexts, Educational Aspirations and Attainment: A Moderation-Mediation Model Extended'. *Learning Environments Research*, 6 (247-265).
Marsh, H. W., Balla, J. R. and McDonald, R. P. (1988). 'Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size'. *Psychological bulletin,* 103 (3), 391-410.
McDonald, R. P. and Marsh, H. W. (1990). 'Choosing a multivariate model: noncentrality and goodness of fit'. *Psychological Bulletin*, 107 (2), 247-255.
Melhuish, E., Romaniuk, H., Sammons, P., Sylva, K., Siraj-Blatchford, I. and Taggart, B. (2006). *The effectiveness of primary schools in England in Key Stage 2 for 2002, 2003 and 2004* (Brief No: RBX06-06). London: Birkbeck University.
Meredith, W. (1993). 'Measurement invariance, factor analysis and factorial invariance'. *Psychometrika*, 58 (4), 525-543.
Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
Millsap, R. E. (2013). personal communication 04/01/2013, 17:20 (see Appendix A7).
Modood, T. (2003). 'Ethnic differentials in educational performance'. In D. Mason (Ed.), *Explaining ethnic differences: changing patterns of disadvantage in Britain*. Bristol The Polity Press.
Modood, T. (2004). 'Capitals, ethnic identity and educational qualifications'. *Cultural Trends,* 13 (2), 87-105.
Modood, T. (Ed.) (2005), *Ethnicity and intergenerational identities and adaptations in Britain*. Cambridge Cambridge University Press.

Modood, T., Berthood, R., Lakey, J., Nazroo, J., Smith, P., Virdee, S. and Beishon, S. (1997). *Ethnic minorities in Britain: Diversity and Disadvantage*. London: Policy Studies Institute.
Muller, D., Judd, C. M. and Yzerbyt, V. Y. (2005). 'When moderation is mediated and mediation is moderated'. *Journal of Personality and Social Psychology,* 89 (6), 852-863.
Murmane, R. J., Maynard, R. A. and Otis, J. C. (1981). 'Home resources and children's achievement'. *The Review of Economics and Statistics* 63 (3), 369-377.
Muthén, B. and Asparouhov, T. (2014). 'Causal effects in mediation modeling: an introduction with applications to latent variables'. *Structural Equation Modeling,* forthcoming.
Muthén, B. and Kaplan, D. (1985). 'A comparison of some methodologies for the factor analysis of non-normal Likert variables'. *British Journal of Mathematical and Statistical Psychology,* 38, 171-189.
Palich, L. E., Horn, P. W. and Griffeth, R. W. (1995). 'Managing in the international context: Testing cultural generality of sources of commitment to multinational enterprises'. *Journal of Management,* 21 (4), 671-690.
Perneger, T. V. (1998). 'What's wrong with Bonferroni adjustments'. *British Medical Journal,* 316 (7139), 1236.
Phillips, D. (1998). 'Black minority ethnic concentration, segregation and dispersal in Britain'. *Urban Studies,* 35 (10), 1681-1702.
Picou, J. S. and Carter, M. (1976). 'Significant-Other Influence and Aspirations'. *Sociology of Education,* 49 (1), 12-22.
Portes, A., McLeod, S. A. J. and Parker, R. N. (1978). 'Immigrant aspirations'. *Sociology of Education,* 51 (4), 241-260.
Qian, Z. and Blair, S. L. (1999). 'Racial/ethnic differences in educational aspirations of high school seniors'. *Sociological Perspectives,* 42 (4), 605-625.
Quaglia, R. J. and Cobb, C. D. (1996). 'Toward a theory of stuident aspirations'. *Journal of Research in Rural Education,* 12 (3), 127-132.
Raykov, T. (2004). 'Behavioral scale reliability and measurement invariance evaluation using latent variable modeling'. *Behavior Therapy,* 35 (2), 299-331.
Raykov, T. and Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.
Riordan, C. M. and Vandenberg, R. J. (1994). 'A central question in cross-cultural research: do employees of different cultures interpret work-related measures in an equivalent manner?'. *Journal of Management,* 20 (3), 643-671.
Ritchie, C., Flouri, F. and Buchanan, A. (2005). *Aspirations and expectations*. Oxford: Centre for Research into Parenting and Children University of Oxford.
Rosen, B. C. and Aneshensel, C. S. (1978). 'Sex differences in the educational-occupational expectation process'. *Social Forces,* 57 (1), 164-186.
Ross, A. and Lloyd, J. (2013). *Access for all: An investigation of young people's attitudes to the cost of higher education using the Longitudinal Study of Young People in England*. London: Strategic Society Centre.
Rothman, K. J. (1990). 'No adjustments are needed for multiple comparisons'. *Epidemiology,* 1, 43–46.
Rothon, C. (2005). The importance of social class in explaining educational attainments of minority ethnic pupils in Britain: evidence fromn the Youth Cohort Study, *Education, Social Mobility and Social Cohesion Workshop (EDUC)*. Mannheim, December 2-3, 2005: Mannheim Centre for European Social Research.

Rumbaut, R. G. (1994). 'The Crucible Within: Ethnic identity, Self-Esteem and Segmented Assimilation Among Children of Immigrants'. *International Migration Review,* 28 (4), 748-794.
Sacker, A., Schoon, I. and Bartley, M. (2002). 'Social inequality in educational achievement and psychosocial adjustment throughout childhood: magnitude and mechanisms'. *Social Science & Medicine,* 55, 863–880.
Schaie, K. W., Maitland, S. B., Willis, S. L. and Intrieri, R. C. (1998). 'Longitudinal invariance of adult psychometric ability factor structures across 7 years'. *Psychology and aging,* 13 (1), 8-20.
Sewell, W. H., Haller, A. O. and Portes, A. (1969). 'The Educational and Early Occupational Attainment Process'. *American Sociological Review*, 34 (1), 82-92.
Sewell, W. H., Hauser, R. M. and Wolf, W. C. (1980). 'Sex, Schooling, and Occupational Status'. *The American Journal of Sociology*, 86 (3), 551-583.
Sewell, W. H. and Shah, P. V. (1968). 'Social Class, Parental Encouragement, and Educational Aspirations'. *The American Journal of Sociology,* 73 (5), 559-572.
Shrout, P. E. and Bolger, N. (2002). 'Mediation in experimental and nonexperimental studies: new procedures and recommendations'. *Psychological Methods*, 7, 422-445.
Siraj-Blatchford, I. (2010). 'Learning in the home and at school: how working class children ‘succeed against the odds’'. *British Educational Research Journal,* 36 (3), 463–482.
Sörbom, D. (1974). 'A general method for studying differences in factor means and factor structure between groups'. *British Journal of Mathematical and Statistical Psychology,* 27, 229-239.
Sörbom, D. (1978). 'An alternative to the methodology for analysis of covariance'. *Psychometrika*, 43, 381-396.
Stacy, A. W., MacKinnon, D. P. and Pentz, M. A. (1993). 'Generality and specificity in health behavior: Application to warning-label and social influence expectancies'. *journal of Applied Psychology,* 78 (4), 611-627.
Steenkamp, J. B. E. M. and Baumgartner, H. (1998). 'Assessing measurement invariance in cross-national consumer research'. *Journal of consumer research,* 25 (1), 78-107.
Steiger, J. H. and Lind, J. C. (1980). 'Statistically-based tests for the number of common factors'. *Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City*.
Stevens, P. A. J. (2007). ''Researching race/ethnicity and educational inequality in English secondary schools: a critical review of the research literature between 1980 and 2005'. *Review of Educational Research,* 77 (147-185).
Strand, S. (1999). 'Ethnic group, sex and economic disadvantage: associations with pupils' educational progress from Baseline to the end of key Stage 1'. *British Educational Research Journal,* 15 (2), 179-202.
Strand, S. (2007). *Minority Ethnic Pupils in the Longitudinal Study of Young People in England (LSYPE)*. (Vol. Research Report No DCSF-RR002). Warwick: University of Warwick.
Strand, S. (2008). *Minority ethnic pupils in the Longitudinal Study of Young People in England: Extention report on performance on national examinations at age 16*. London: DCSF.
Strand, S. and Winston, J. (2008). 'Educational aspirations in inner city schools'. *University of Warwick institutional repository: http://go.warwick.ac.uk/wrap*.
Stryker, S. (2007). 'Identity theory and personality theory: Mutual relevance'. *Journal of Personality* 75 (6), 1083-1102.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2004). *The Effective Provision of Pre-School Education [EPPE] Project: A Longitudinal Study funded by the DfES1997 – 2004* (Final Report). London: Institute of Education.
Tabachnick, B. G. and Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn and Bacon.
Tanaka, J. K. (1993). 'Multifaceted conceptions of fit in structural equation models'. In K. A. Bollen and S. J. Long (Eds), *Testing structural equation models*. Newbury Park, CA Sage Publications Inc.
Thompson, M. S., Alexander, K. L. and Entwisle, D. R. (1988). 'Household composition, parental expectations and school achievement'. *Social Forces*, 67 (2), 424-451.
Tyson, E. H. (2004). 'Ethnic differences using behavior rating scales to assess the mental health of children: A conceptual and psychometric critique'. *Child psychiatry and human development*, 34 (3), 167-201.
Tzanakis, M. (2014). *An exploration of ethnic differences of the influences of home and school on the development of young people's educational expectations for university study between ages 14 to 16*. Unpublished Doctoral Thesis. Institute of Education, University of London, London.
Vandenberg, R. J. and Lance, C. E. (2000). 'A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research'. *Organizational research methods,* 3 (1), 4-70.
Wicherts, J. M., Dolan, C. V. and Hessen, D. J. (2005). 'Stereotype threat and group differences in test performance: a question of measurement invariance'. *Journal of Personality and Social Psychology,* 89 (5), 696-716.
Widaman, K. F., Ferrer, E. and Conger, R. D. (2010). 'Factorial invariance within longitudinal structural equation models: Measuring the same construct across time'. *Child development perspectives,* 4 (1), 10-18.
Widaman, K. F. and Reise, S. P. (1997). 'Exploring the measurement invariance of psychological instruments: Applications in the substance use domain', *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Unversity of Houston, Texas: http://people.cehd.tamu.edu/~okwok/epsy651R/Articles/WidamanReise1997.pdf (last accessed: 3-1-2013).
Wilson, D., Burgess, S. and Briggs, A. (2005). The dynamics of school attainment of England's ethnic minorities, *Centre for Market and Public Organisation* Bristol: University of Bristol, Department of Economics.
Winkel, G., Saegert, S. and Evans, G. W. (2009). 'An ecological perspective on theory, methods, and analysis in environmental psychology: Advances and challenges'. *Journal of Environmental Psychology*, 29 (3), 318-328.
Woelfel, J. and Haller, A. O. (1971). 'Significant others, the self-reflexive act and the attitude formation process'. *American Sociological Review,* 36 (1), 74-87.