

DRAFT PAPER. COMMENTS WELCOME.  
PLEASE DO NOT QUOTE WITHOUT AUTHOR'S PERMISSION.

Paper presented at the British Educational Research Association Annual Conference,  
University of Glamorgan, 14-17 September 2005

## **SUMMARY STATISTICS, EDUCATIONAL ACHIEVEMENT GAPS AND THE ECOLOGICAL FALLACY**

**Paul Connolly**  
**Queen's University Belfast, UK<sup>1</sup>**

### **ABSTRACT**

*Summary statistics continue to play an important role in identifying and monitoring general trends in relation to patterns of educational inequalities between differing groups of pupils over time. However, this paper argues that their uncritical use can also encourage the labelling of whole groups of pupils as 'underachievers' or 'overachievers' as the findings of group-level data are simply applied to individual group members, a practice commonly termed the 'ecological fallacy'. Some of the adverse consequences of this will be outlined in relation to current debates concerning gender and ethnic differences in educational attainment. It will be argued that one way of countering this uncritical use of summary statistics and the ecological fallacy that it tends to encourage, is to make much more use of the principles and methods of what has been termed 'exploratory data analysis'. Such an approach is illustrated through a secondary analysis of data from the Youth Cohort Study of England and Wales, focusing on gender and ethnic differences in educational attainment. It will be shown that, by placing an emphasis on the graphical display of data and on encouraging researchers to describe those data more qualitatively, such an approach represents an essential addition to the use of simple summary statistics and helps to avoid the limitations associated with them.*

### **INTRODUCTION**

This paper is about how the nature and extent of inequalities in educational attainment might be most appropriately described and analysed using quantitative data. That this is an issue of significant political importance can be illustrated by the ongoing moral panic surrounding boy's underachievement in education. Since the mid-1990s, concerns over the apparent tendency for boys to be lagging further and further behind girls in public examinations have tended to dominate educational debates with the consequences of not only leading to the skewing of efforts and resources towards tackling this 'problem' but

---

<sup>1</sup> Professor Paul Connolly, School of Education, Queen's University Belfast, 69/71 University Street, Belfast BT7 1HL, Northern Ireland. Tel. +44 (0)28 9097 5952. Fax. +44 (0)28 9023 9263. Email. paul.connolly@qub.ac.uk

also to the marginalisation of concerns over more significant and enduring inequalities in educational attainment, most notably those surrounding social class and ethnicity (Yates, 1997; Epstein *et al.*, 1998a; Raphael Reed, 1999). At the heart of this moral panic lies the tendency to make use of crude summary statistics that simply compare the average performance of boys with girls. This, in turn, has tended to lead to crude generalisations being made about *all* boys and *all* girls as if they represent two distinct and homogenous groups (Mac an Ghail, 1996; Jackson, 1998). As a number of commentators have pointed out, what this tendency to take simple group-level data and apply it uncritically to all individuals within that group – a tendency termed the ecological fallacy – tends to ignore is the considerable variation that exists within each group and also the significant overlap in levels of performance between boys and girls. Thus not all boys are underachieving and, similarly, not all girls are performing well (Epstein *et al.*, 1998b; Lucey and Walkerdine, 2000; Connolly, 2004).

This tendency for summary statistics to not only construct distorted and misleading pictures but also to result in adverse political effects is not solely associated with gender differences however but has also been found in relation to ethnicity as well. As far back as the publication of the Rampton Report (1981), concerns were being raised regarding the way in which the crude comparisons that it made between the educational attainment of ‘White’, ‘Asian’ and ‘Black’ students tended to label all Black students as ‘underachievers’ (Troyna, 1984). Moreover, this tendency towards committing the ecological fallacy had important political consequences in tending to individualise the problem and thus detracting attention away from what was considered by some to be the real issue at hand - institutional racism. As Troyna (1984: 154) argued:

Distorted and grossly over-simplified empirical studies [of ethnic differences in educational attainment] have not only been highly damaging to the way educationists, and practitioners particularly, perceive the intellectual capabilities of black pupils, but also have led directly to the development of educational policies and practices which are largely irrelevant to an enhancement of the life chances of these pupils.

More recent research has tended to reinforce these concerns over the use of crude statistical summaries to compare the performance of different ethnic groups. It has been shown, for example, that there are not only important distinctions to be made within the Black and Asian categories previously used but also that there is considerable variability at a local level (Gillborn and Gipps, 1996; Gillborn and Mirza, 2000). As regards the latter point, Gillborn and Mirza (2000) found in their study of data on ethnic differences in educational performance produced by local education authorities (LEAs) in England and Wales that, of the six minority ethnic categories they analysed, every one was found to be the highest attaining group in at least one LEA. Overall, however, such a critique has only been partially successful. While it has tended to move debates on from those based simply around the notion of ‘Black underachievement’ they have tended simply to be replaced with a concern over what is now perceived to be an ‘underachieving cluster’ of Black, Pakistani and Bangladeshi pupils that is falling further behind the others (Demack *et al.*, 2000). The potential for such analyses to continue to encourage the

ecological fallacy and thus the labelling of whole groups of pupils as ‘underachievers’ is therefore still clearly evident (Haque, 2000).

It is with these concerns in mind that this paper argues for the need to compliment the use of simple summary statistics with additional methods of describing and analysing quantitative data that can fully illustrate the complexities that exist in relation to patterns of achievement and thus help to counter the potential to commit the ecological fallacy. Through a secondary analysis of data from the Youth Cohort Study of England and Wales, the nature of the problems associated with the uncritical use of summary statistics to describe and analyse gender and ethnic differences are outlined in a little more detail. The paper then goes on to show how these problems can be effectively addressed by an approach based upon the principles and methods of what has been termed ‘exploratory data analysis’, as espoused most notably by Tukey (1977), that places an emphasis on contextualising the use of summary statistics with the graphical display of data and on encouraging the researcher to describe these data a little more qualitatively. Before moving onto the substantive analysis, however, it is necessary to provide some details on the Youth Cohort Study for England and Wales and the nature of the dataset and the measures of educational attainment to be used here.

## METHODOLOGY

The Youth Cohort Study of England and Wales is the largest nationally-representative longitudinal survey of 16-19 year olds in the UK (Demack *et al.*, 2000). It is also the only one of its type that includes details on young people’s educational attainment and their gender, ethnicity and social class background. Since 1985 there have been 11 cohorts of young people first contacted shortly after they reached minimum school leaving age and who have then been tracked over the following three years to follow their progress in the educational system and/or labour market. The data upon which the analysis in this paper is based is derived from the first sweep of the tenth cohort of young people to be followed that represents the latest cohort for which data has been made publicly available for secondary analysis.<sup>2</sup> These young people finished compulsory schooling during the summer of 1999. The population from which they were sampled consisted of all males and females in England and Wales who had reached the age of 16 on 31 August 1999. A random sample was then selected from this population via schools that initially totalled 25,000 young people. For this first sweep, those selected were sent a questionnaire to complete in February 2000 and this was followed up by a number of reminders and finally an attempt to contact those who had still not responded by telephone. Overall, a final sample of 13,698 was achieved that represented a response rate of 55%.

Given this relatively low response rate, a weighting system was devised and applied to the resulting dataset to correct for any known biases that may have occurred due to non-response. The weights were constructed using four variables that were known to be associated with non-response: gender, qualifications achieved, region that the respondent lived in and the type of school attended (for further details see RSGB, 2001). All of the analysis to follow is based upon the weighted dataset.<sup>3</sup> While the application of weights

---

<sup>2</sup> See <<http://www.data-archive.ac.uk>> for further details.

<sup>3</sup> One issue to clarify is whether weighting the sample partly in relation to gender will then impact upon the consequent analysis of gender differences. In essence, if a dataset is weighted simply in terms of gender

may well have partially corrected for some of the biases introduced by the low response rate it is unlikely to have completely addressed the problem. It is therefore important to bear this limitation in mind when considering the findings below. The main characteristics of the final (weighted) sample are shown in Table 1.

The first questionnaire sent out to young people covered a range of issues including the educational qualifications achieved by respondents as well as their immediate educational and career choices and destinations for the year after leaving school. The data upon which this paper is based focus specifically on the respondents' educational

Table 1. Breakdown of the Sample by Gender, Social Class and Ethnicity (n=13,698)\*

|                         | n      | %     |
|-------------------------|--------|-------|
| <i>Gender</i>           |        |       |
| Males                   | 6,938  | 50.6  |
| Females                 | 6,760  | 49.4  |
|                         | 13,698 | 100.0 |
| <i>Social Class</i>     |        |       |
| Professional/Managerial | 2,786  | 20.3  |
| Other Non-Manual        | 2,545  | 18.6  |
| Skilled Manual          | 4,603  | 33.6  |
| Semi-Skilled Manual     | 1,449  | 10.6  |
| Unskilled Manual        | 597    | 4.4   |
| Other                   | 40     | 0.3   |
| Not Classified          | 1,677  | 12.2  |
|                         | 13,697 | 100.0 |
| <i>Ethnicity</i>        |        |       |
| White                   | 11,768 | 85.9  |
| Indian                  | 360    | 2.6   |
| Pakistani               | 342    | 2.5   |
| Black                   | 331    | 2.4   |
| Bangladeshi             | 134    | 1.0   |
| Chinese                 | 78     | 0.6   |
| Other                   | 280    | 2.0   |
| Not Answered            | 404    | 2.9   |
|                         | 13,697 | 99.9  |

\*Some of the frequencies may not sum to 13,698 due to rounding following the application of weights. Also, percentages may not sum to 100.0 due to rounding.

then this will effect the overall estimates produced in relation to the total numbers and thus proportions of males and females in the sample. The main reason for doing this is simply to produce estimates for the population that are less biased and thus more representative. However, more specific estimates in relation to the proportion of boys (or girls) achieving five or more GCSE Grades A\*-C or the mean GCSE Scores (and their corresponding standard deviations) for boys (or girls) will remain unaffected.

attainment at the end of compulsory schooling (i.e. the end of Year 11). Two measures of attainment are used below, the first is simply the proportions of respondents who gained five or more GCSEs (General Certificates of Secondary Education) grades A\*-C – to be referred to in the following as the ‘GCSE Benchmark’. This is perhaps the most commonly used and understood measure of attainment at this level and is a key benchmark used by government in the derivation of school leagues tables. The second measure is the ‘GCSE Score’ for each respondent. This is calculated simply by allocating points for each grade obtained (i.e. in this case: A\*/A=7, B=6, C=5, ... G=1) and then adding together the points achieved in relation to all of the GCSEs taken. Thus, a respondent who gained five C grades would have achieved an overall GCSE Score of 25 (i.e. 5x5) whereas someone who achieved just two D grades and a G would have a Score of 9 (i.e. 2x4 + 1x1).

One complicating factor in all of this is the fact that the government have for some years included equivalent vocational qualifications – particularly General National Vocational Qualifications (GNVQs) – in their calculations of the GCSE Benchmark. Simply for the sake of consistency the two measures to be used below – GCSE Benchmark and GCSE Score – have also been calculated to include GNVQ equivalents. The means by which equivalent GCSE points have been calculated from GNVQs is outlined in Table 2.

Table 2. Equivalent GCSE Point Scores for GNVQ Qualifications Achieved\*

| Qualification | Level        | Grade       | GCSE Equivalent Points |
|---------------|--------------|-------------|------------------------|
| GNVQ Part One | Intermediate | Distinction | 15                     |
|               |              | Merit       | 12                     |
|               |              | Pass        | 10                     |
|               | Foundation   | Distinction | 8                      |
|               |              | Merit       | 6                      |
|               |              | Pass        | 3                      |
| GNVQ Full     | Intermediate | Distinction | 30                     |
|               |              | Merit       | 24                     |
|               |              | Pass        | 20                     |
|               | Foundation   | Distinction | 16                     |
|               |              | Merit       | 12                     |
|               |              | Pass        | 6                      |

\*Source: DfES (2002: 7)

## THE DISCOURSE OF ‘ACHIEVEMENT GAPS’

Simple summary statistics regarding gender and ethnic differences in educational attainment using these two measures in relation to the sample are provided in Table 3. The first thing to notice about these statistics, when presented in isolation in this way, is the way in which they tend to reduce the performance of whole categories of pupils to single summary points. Lost is any real and immediate sense of the variability of scores within groups and/or the way in which the performance of those within one group will tend to considerably overlap those of another. Reported uncritically and in isolation like this, it is not too surprising to see how the ecological fallacy can emerge. Taking gender differences in the GCSE Benchmark as an example, it can be seen that the performance of girls has been reduced to the single summary point of 53.8% and boys to 43.9%. The focus for attention is then often drawn to the ‘gap’ (of 9.9 percentage points in this case) that exists between the two groups. While the analysis of such overall differences between group averages and particularly how they tend to change over time remains important, if presented simply like this it can easily lead to slippages in the description of the data as showing how ‘boys are lagging behind girls’ which, with just one more step, can fuel the ecological fallacy by implying that it is actually *all* boys lagging behind *all* girls.

In fact this is precisely how the media within the UK has tended to report gender differences in educational attainment since the mid-1990s (Arnot *et al.*, 1999), often by focusing solely on the ‘gap’ and invoking the imagery and metaphor of a race between these two homogenous ‘teams’. Following the publication of the GCSE results in August 2004, for example, a headline in *The Times* (26 August) claimed that ‘Boys are Narrowing the Gender Gap’ and the accompanying article reported how: ‘After trailing ever more distantly behind their female classmates, boys have improved their GCSE results by more than the girls for the second successive year’. A similar picture could be

Table 3. Summary Statistics on the Educational Attainment of School Leavers in England and Wales in 1999 by Gender and Ethnicity

|                  | % Gaining 5 or More GCSEs Grades A*-C | Average GCSE Score |                    |
|------------------|---------------------------------------|--------------------|--------------------|
|                  |                                       | Mean               | Standard Deviation |
| <i>Gender</i>    |                                       |                    |                    |
| Girls            | 53.8                                  | 41.6               | 19.0               |
| Boys             | 43.9                                  | 37.1               | 19.1               |
| <i>Ethnicity</i> |                                       |                    |                    |
| Chinese          | 73.4                                  | 46.7               | 22.6               |
| Indian           | 60.3                                  | 45.2               | 17.8               |
| White            | 49.9                                  | 39.8               | 18.9               |
| Black            | 38.7                                  | 34.3               | 18.7               |
| Pakistani        | 29.4                                  | 33.1               | 17.6               |
| Bangladeshi      | 28.4                                  | 34.0               | 18.6               |

found in *The Guardian* (26 August) where, under the headline ‘GCSE Pass Rate Falls’, it was reported that: ‘Girls continue to outperform boys at this age, though the gap narrowed slightly’. This imagery of a race was even more explicit in the *Times Educational Supplement* (27 August) in its coverage of GCSE results in Wales for that same year. As an almost mirror-image of the trends in England, the headline reported that ‘GCSE Pass Rates Rise’, and the accompanying article noted how: ‘Welsh pupils continue to achieve better results at GCSE than their English counterparts, but high-achieving girls are leaving boys even further behind in the exams race’.

Clearly, one of the problems associated with the descriptions of gender differences outlined above is the rather crude and blunt measure of educational attainment used. It could thus be argued that it is not the use of summary statistics that is at fault *per se* but the GCSE Benchmark itself that tends to draw artificial lines in terms of attainment levels and thus encourages such simplistic comparisons. It is certainly the case that the summary statistics provided in Table 3 for the more sensitive alternate measure of educational attainment – the GCSE Score – have the potential to avoid this construction of differences in terms of a clear ‘gap’. More specifically, the standard deviations given alongside the mean GCSE Scores for boys and girls should provide the reader with a sense of the range of scores obtained by both groups and thus the considerable overlap that exists between them.<sup>4</sup>

There are three problems with this argument, however. The first is that it relies upon the assumption that most readers understand what standard deviations are and will thus be able to interpret these figures accordingly. Unfortunately, this is often not a safe assumption to make (Gorard, 2001). Secondly, the lack of a graphical display of these data and an emphasis simply on reporting the means and standard deviations can so easily still encourage the reader to reduce boys and girls to two summary points (in the case of GCSE Scores of 37.1 and 41.6 respectively) and thus to think of these differences in terms of an actual gap (in this case 4.5 points). Third, and following on from this last point, the problem actually remains as much with the over-reliance on summary statistics as with the particular measure of attainment used. Even in the case of the GCSE Benchmark, for example, it will be shown below that it is possible to present the data in such a way as to avoid (or at least notably reduce) the tendency to generate a discourse of achievement gaps.

Very similar points can be made concerning the summary statistics regarding ethnic differences as also shown in Table 3. As can be seen, whether one takes the GCSE Benchmark or GCSE Score measure, respondents in each case have again been reduced to just six summary points representing the six main ethnic groups. Again, such summary statistics are extremely important and useful in drawing attention to racial inequalities in education. However, if presented in isolation like this they can also encourage the

---

<sup>4</sup> For those who may not be familiar with the notion of the standard deviation, it is basically a measure of the spread of scores. It is estimated that just over two thirds (68%) of scores will be within the range of one standard deviation plus or minus the mean. Thus, in relation to girls in the present study, the standard deviation of 19.0 indicates that two thirds of girls have achieved GCSE Scores within the range  $41.6 \pm 19.0$  i.e. between 22.6 (i.e.  $41.6 - 19.0$ ) and 60.6 (i.e.  $41.6 + 19.0$ ). From the same calculations it can also be seen that approximately two thirds of boys have also obtained GCSE Scores in the range 18.0 to 56.2. This therefore should indicate to the reader that considerable overlap exists in the respective scores of the two groups.

ecological fallacy once more with the tendency to rank the different ethnic groups in terms of overall achievement levels. It is only one further step from this to beginning to describe the findings in terms of Chinese pupils being 'out in front' followed by Indian and White pupils with what would appear to be an 'underachieving cluster' of Pakistani and Bangladeshi pupils at the rear (and possible Black pupils if one focuses on GCSE Scores).

This tendency to describe the educational attainment of minority ethnic groups in this way can be fuelled through the use of what are known as inferential statistics. The main aim of such statistics is to indicate the degree of accuracy with which estimates from a randomly-selected sample can be taken as representative of the population as a whole.<sup>5</sup> One such method, as used by Demack *et al.* (2000) in their analysis of ethnic differences in the mean GCSE Scores between groups, is the use of error bars. This is illustrated in Figure 1 that represents the mean GCSE Scores for the six ethnic groups comprising the current sample and broadly replicates the approach used by Demack *et al.* (2000: 128) previously. As can be seen, the dots represent the mean GCSE Score for each respective group and tally with those provided earlier in Table 3. The error bars protruding from these illustrate what is known as the 95 per cent confidence interval. In other words, as these figures are only estimates of the GCSE Scores of each ethnic group taken from a random sample, they are likely to include some error due simply to the vagaries of sample selection. What these bars indicate is the extent of that error. Thus, in this case, we can be 95 per cent confident that the actual mean GCSE Score lies somewhere between the limits of these error bars. It is interesting to note that the error bars are smallest for White pupils and much larger for the rest. This simply reflects the much larger sub-sample used to produce the estimate of the mean GCSE Score for White pupils and thus the much reduced level of error that is consequently likely to accompany this.

As before, such summary findings are important in helping us to interpret the data and assess the extent to which we can generalise the findings derived from these to the population as a whole. In relation to Figure 1, for example, it can be seen that the mean GCSE Scores for the Black, Bangladeshi and Pakistani groups do differ slightly. However, the overlapping error bars suggest that the potential errors caused by the relatively small sub-sample sizes used mean that we cannot be confident that these differences reflect real differences in the population as a whole, rather than random error.

---

<sup>5</sup> It is sometimes argued that with such a low response rate of just 55% then there is little point using inferential statistics to analyse the data given that any error due to random sampling that such an analysis would help identify is likely to be completely overshadowed by the errors resulting from non-response. However, a more pragmatic approach is to make use of as much information as possible. Clearly, a key concern should be to assess what potential biases have been introduced by non-response and to correct for these as best one can. However, and beyond this, inferential statistics can still play a role in aiding the interpretation of any findings derived from the data. If it is shown with the use of inferential statistics, for example, that a finding could have occurred simply due to the random nature of the sample selected rather than being likely to reflect something real within the population as a whole then this is useful information. Regardless of the additional biases resulting from non-response, this would be enough in itself to discount the finding. However, the low response rate would mean that we could not simply accept uncritically any findings that were sufficiently confirmed through inferential statistics. While in such circumstances we would be able to conclude that our finding is likely not to have simply been caused by sampling error we would still need to be mindful of the possible errors introduced by non response.

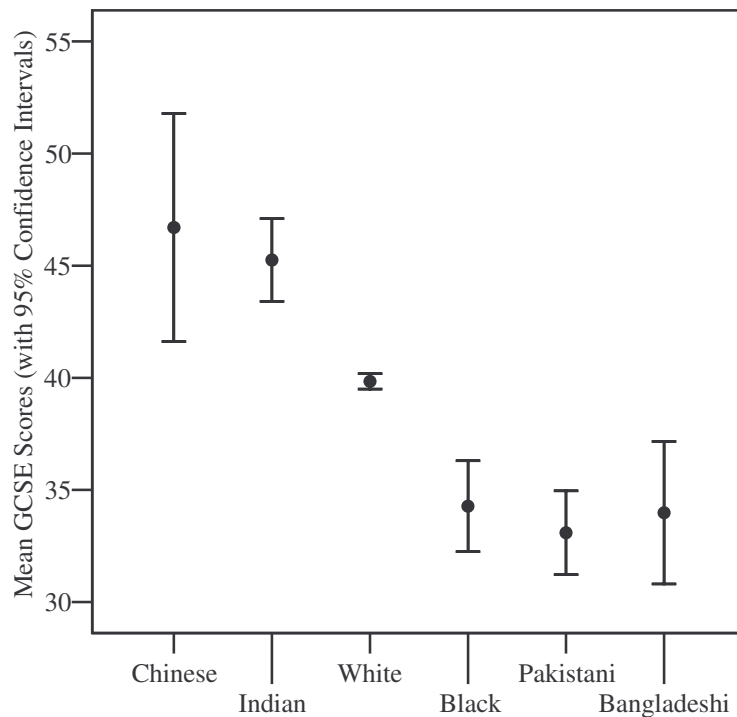


Figure 1. Ethnic Differences in Mean GCSE Scores Attained by School Leavers in England and Wales in 1999

Thus, one conclusion we can draw from Figure 1 is that while we may have found differences in the mean GCSE Scores between all six ethnic groups, we simply do not have sufficient evidence that the differences between Chinese and Indian pupils and also the differences between Black, Bangladeshi and Pakistani pupils are real. However, because of the lack of overlap between some of the error bars we can conclude that there is sufficient evidence that the mean GCSE Score for White pupils is higher than for Black, Bangladeshi and Pakistani pupils. Similarly, we can also conclude that there is sufficient evidence that the mean scores for Chinese and Indian pupils are higher than for White pupils.

However, while this information is extremely useful there are two main potential problems when it is presented in isolation in this way. The first is that it provides no opportunity to challenge the tendency to reduce whole groups of pupils to simple summary points and thus it has the potential to continue to encourage the ecological fallacy. In relation to Figure 1, for example, there is still no information provided on the range and spread of GCSE Scores and thus no indication given of the overlap in scores between groups.

Second, beyond assessing how well these findings can be generalised to the population as a whole, they are actually of very limited use practically in determining whether there is actually an 'underachieving cluster' even simply in terms of this group-level of analysis. This is because, as already indicated above, any findings regarding differences between the mean scores of groups are wholly dependent upon the sample

sizes used. For example, and as already mentioned, the estimated mean GCSE Scores for Black, Bangladeshi and Pakistani pupils are not the same but do differ to a certain extent. If we increased all of the sub-samples ten-fold then the estimates of the GCSE Scores would be so accurate, and thus the error bars so small, that there would be no overlapping and thus no clustering at all. Similarly, if we used greatly reduced sub-samples then we can presume that the potential errors would be much greater and thus we would have completely overlapping error bars suggesting that we have no evidence at all of any ethnic differences in mean GCSE Scores.

Overall, the key point is that inferential statistics of the type illustrated in Figure 1 have a very specific use in helping us understand the level of error associated with estimates derived from a random sample. While this is important in itself, there is a danger of then using such inferential analysis inappropriately in ways that can potentially provide misleading representations of differences between groups and thus encourage further the ecological fallacy.

### **EXPLORATORY DATA ANALYSIS (EDA)**

The dangers associated with relying solely on summary statistics when describing and analysing inequalities in educational attainment should be clear by now. However, it is worth emphasising the points made above further by looking again at the data but this time using some of the principles and methods of exploratory data analysis (EDA). In doing so, this will also serve as an illustration of the benefits of using EDA to help contextualise the summary statistics that are produced and thus to effectively counter the potential dangers of the ecological fallacy associated with these. EDA itself represents not just a wide range of techniques for analysing quantitative data but also an underlying philosophy and was first introduced by Tukey (1977). It can be understood partly as a response to concerns with the way in which quantitative data analysis had become equated simply with statistics and thus the use of statistical summaries and of significance testing (Hartwig and Dearing, 1979). For Tukey (1977), EDA should be seen as detective work with an emphasis being placed on gaining as much information about the data and how it is distributed as possible. This, in turn, requires the use of graphical methods to display the data in as many different ways and formats as possible so as to gain a true feel for what is going on and also to see the unexpected. With this in mind, and as Hartwig and Dearing (1979: 9) contend:

One should be sceptical of measures which summarise data since they can sometimes conceal or even misrepresent what may be the most informative aspects of the data, and one should be open to unanticipated patterns in the data since they can be the most revealing outcomes of the analysis.

While summary statistics are not dismissed as such, an EDA approach has tended to emphasise the necessity of understanding the data and what is to be summarised first before then generating appropriate (and resistant) summary measures (Hoaglin, Mosteller and Tukey, 1983). Moreover, and emanating from this, there is a concern with the extremely limiting nature of significance testing that, for proponents of EDA, seems to

have become the dominant mode of quantitative data analysis. As Hartwig and Dearing (1979: 10) continue:

In this confirmatory model of analysis, a model for the relationship (often linear) is fitted to the data, statistical summaries (such as means or explained variances) are obtained, and these are tested against the probability that values as high as those obtained could have occurred by chance. Not only does this mode of analysis place too much trust in statistical summaries but it also lacks openness since only two alternatives are considered. The data are not explored to see what other patterns might exist.

It is impossible in such a short space to do justice to the richness and detail that is to be found in EDA. Moreover, it is not being claimed here that the approach below is strictly EDA but, rather, that it is simply informed by it. Nevertheless, by looking briefly again at the same data discussed above and placing an emphasis on graphical representations of the data, the benefits of EDA should soon be apparent.

To begin with gender and educational attainment, the data relating to differences between boys and girls in relation to their GCSE Scores can be illustrated by means of a simple back-to-back histogram as shown in Figure 2. As can be seen what is immediately apparent are the *similarities* in levels of attainment between boys and girls rather than the differences. On closer inspection there is a slight skewing of the respective distributions evident with more girls gaining higher scores than boys but this is only relatively small. Perhaps the most important point to make is that the ‘gap’ that exists when conducting analysis at the level of the group is not evident at the level of the individual. When comparing individual boys and girls therefore there simply is no gap and thus it makes no sense to speak in terms of boys lagging behind girls.

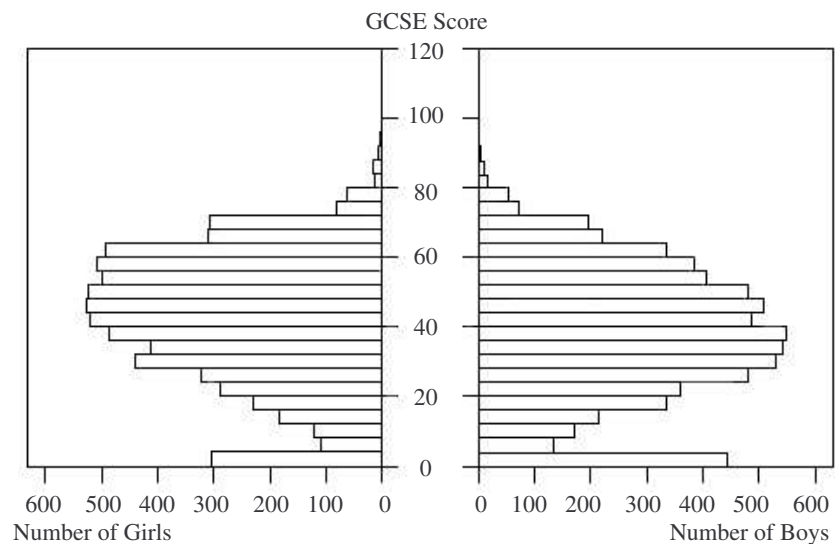


Figure 2. Back-to-Back Histogram Illustrating Gender Differences in GCSE Scores attained by School Leavers in England and Wales in 1999

One of the benefits of displaying all of the data in this way is that it requires the researcher to identify and describe the key features of these more qualitatively. In doing this, and as seen above, it prevents the researcher from committing the ecological fallacy of believing that there is an actual ‘gender gap’. This is also true for the alternative measure of educational attainment, the GCSE Benchmark. As can be seen in Figure 3, the emphasis here is on displaying *all* of the data not just parts of it as often happens with the use of bar charts. In this latter case the bar chart would often have two vertical bars just showing the proportions achieving the GCSE Benchmark (53.8% for girls, 43.9% for boys). The reader is then immediately drawn to compare the differing heights of the two bars and this can, in turn, encourage the ecological fallacy of seeing girls’ levels of achievement as being higher than boys. However, a different picture emerges when displaying all of the data as in Figure 3. In this case the considerable overlap between the attainments of girls and boys is evident. In fact the vast majority of boys and girls (90.1%) can be ‘matched-up’ in terms of sharing the same level of achievement as categorised by this admittedly crude measure of attainment.<sup>6</sup> Only in a small minority of cases (9.9% to be precise<sup>7</sup>) are girls achieving more than boys. Again, while there is a difference it can be seen that it is relatively small and that there is certainly no ‘gap’

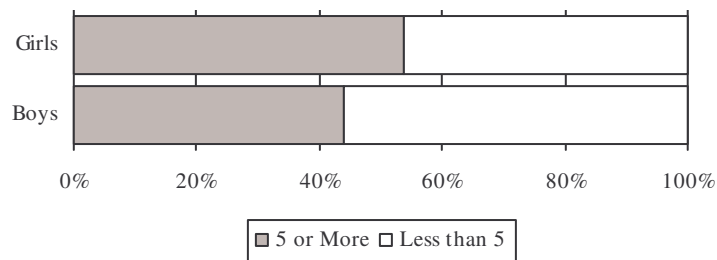


Figure 3. Proportions of School Leavers in England and Wales Gaining 5 or More GCSEs Grades A\*-C in 1999 by Gender

<sup>6</sup> Clearly, it could be argued that this is a simplistic and misleading way to interpret the data given that this is such a crude measure of attainment. More specifically, it could be argued that simply because two pupils are categorised as ‘achieving five or more GCSE Grades A\*-C’ this does not mean they can be ‘matched up’ as suggested here. For example, one may just have five GCSE C grades while the other may have ten GCSE A\* grades. However, to argue this is rather missing the point. Presumably by this logic, if both students gained ten GCSE A\* grades then they could be ‘matched up’. However we could then go a stage further. It may be that the first student only just gained the required marks in each GCSE to achieve an A\* grade while the other student achieved almost perfect scores each time. The key point therefore is that any measure of educational attainment is arbitrary. When examining gender or ethnic differences as we are doing in this article, all we can do is to use the measures that are provided. In relation to the GCSE Benchmark therefore educational attainment is simply measured in terms of whether a pupil gains five or more GCSE Grades A\*-C. By this measure we can ‘match up’ 91.1% of boys and girls. Whether this is an appropriate and valid measure of educational attainment is another matter.

<sup>7</sup> In other words, the ‘extra’ 9.9% of girls achieving the GCSE Benchmark compared to boys (i.e. from the figures in Table 3 this is calculated as: 53.8 – 43.9). It will be noted that this is what is commonly referred to as the ‘gender gap’ as discussed earlier. However, it should be apparent from this discussion that it makes no sense to see it in terms of a gap as such as this implies a real distance (of 9.9 percentage points) between all boys and all girls.

separating the two groups of pupils.

This same approach can also be taken in terms of examining ethnic differences in the GCSE Benchmark as illustrated in Figure 4. Two key points are evident from this. First, when comparing this figure with Figure 3 it is clear that ethnicity is much more strongly associated with the educational attainment of pupils than gender is. If we compare Chinese and Bangladeshi pupils, for example, then it can be seen that just under half (45% to be precise) of the Chinese pupils are achieving higher levels of attainment than the Bangladeshi pupils.<sup>8</sup> However, while ethnic differences like this should be a cause for concern it is also misleading to describe these in terms of a ‘gap’ as such. More specifically, another way to look at the data is that just over half of Chinese and Bangladeshi pupils (55%) can be ‘matched-up’ in terms of sharing the same levels of attainment. Seeing it in this way helps to counter the ecological fallacy with its tendency to label all Bangladeshi pupils, in this case, as ‘underachievers’ (Haque, 2000).

Of course the GCSE Benchmark is a rather crude and blunt measure that tends to artificially exacerbate the differences in educational attainment that exist. If one examines the differences between ethnic groups in terms of their GCSE Score then the amount of overlap between groups and the dangers of applying labels to some is even more apparent. Unfortunately, because there are six ethnic groups here it is not possible to use the simple back-to-back histogram chart used in relation to gender differences in GCSE Scores (Figure 1). However, Tukey (1977) has developed a range of graphical means for displaying data precisely to address this problem and to allow more complex comparisons. One such method is what is now commonly termed the ‘boxplot’. As this may be new to some then it is worth briefly explaining how to interpret the boxplot before then using these to illustrate ethnic differences in GCSE Scores.

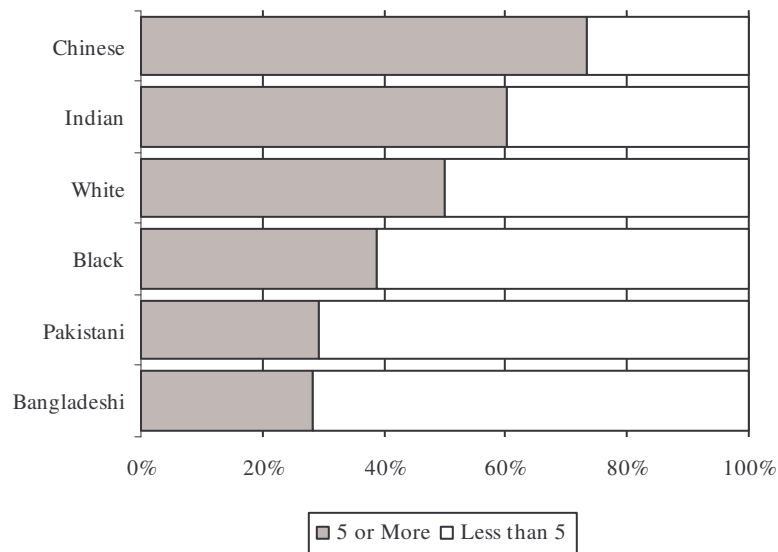


Figure 4. Proportions of School Leavers in England and Wales Gaining 5 or More GCSEs Grades A\*-C in 1999 by Ethnicity

<sup>8</sup> As with the gender differences discussed earlier, this is the ‘extra’ proportion of Chinese pupils achieving the GCSE Benchmark compared to Bangladeshi pupils (i.e. from 73.4 – 28.4).

By way of illustration, and as shown in Figure 5, boxplots have been added to the original back-to-back histogram for gender differences provided in Figure 2. As can be seen, there is one boxplot for girls and one for boys. Both of them represent the distribution of GCSE Scores for each of the two groups respectively. To explain the essential features of the two boxplots, the horizontal line in the middle of each box represents the median score for girls and boys respectively. This is the ‘middle’ of the distribution such that precisely half of the girls (or boys) have gained GCSE Scores above this line and half below. The actual median score for girls is 43 and for boys is 37. The ‘box’ itself represents what is called the inter-quartile range. In simple terms, this means that precisely half of all girls (or boys) will have GCSE Scores within the range indicated by this box. Moreover, a further quarter of girls (or boys) will have scores below the lower edge of the box and a quarter above the upper edge of the box. The ends of each line represent the lowest and highest scores respectively (not counting outliers and extreme case). It can therefore be read off from the boxplot that the lowest GCSE Score obtained for girls was 0 and the highest 96 whereas, for boys, it was 0 and 91 respectively. Finally, the ‘outliers’ are marked individually as circles and represent those individual girls (or boys) who gained GCSE Scores deemed to be well above or below what was expected for that group.<sup>9</sup> It can be seen, for example, that three girls gained exceptionally high scores (98, 103 and 105 respectively) whereas four boys did (scores of 93, 98, 101, 105 and 112 respectively).

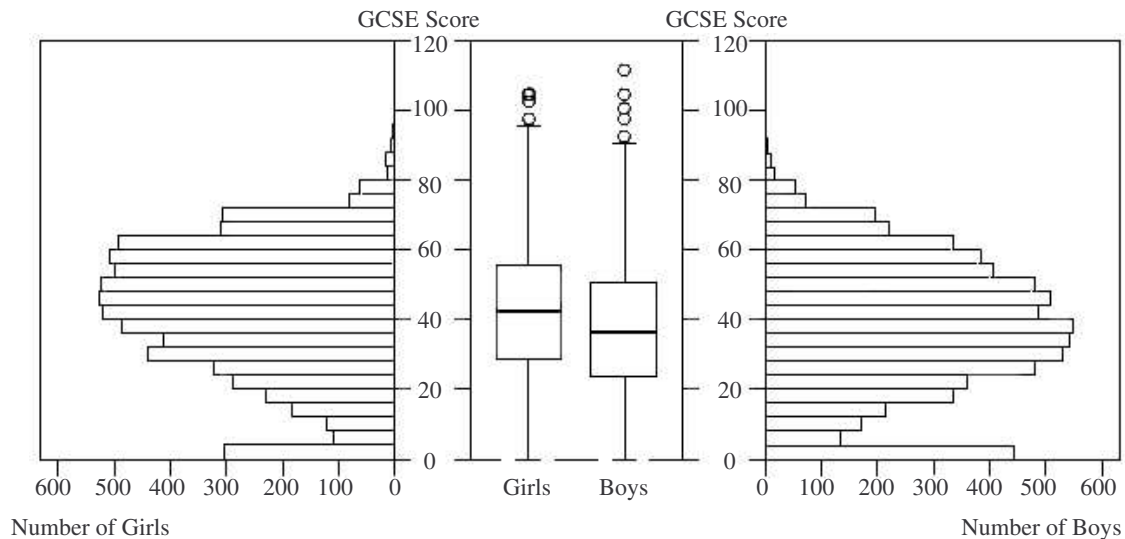


Figure 5. Back-to-Back Histogram with Boxplots Superimposed Illustrating Gender Differences in GCSE Scores Attained by School Leavers in England and Wales in 1999

<sup>9</sup> There are no universal criteria used for deciding where the ‘cut-off’ point is in terms of designating individual scores as ‘outliers’ or not (Hartwig and Dearing, 1979). SPSS calculates outliers as being beyond two inter-quartile ranges from the median. Thus, for girls, their median score is 43 and the inter-quartile range is calculated as 27 (this, it will be remembered, is also the distance between the upper and lower edges of the box). Outliers are therefore classified as those  $\pm 54$  from this (i.e.  $2 \times 27$ ). Thus, in terms of upper scores this would mean anything above 97.

The outliers do not tend to show up on the histograms simply because of the scales used. However, putting the boxplots against their respective histograms as done in Figure 5 should still illustrate what the boxplot is aiming to represent and how it is taken from the histogram. Overall, it can be seen that putting boxplots side-by-side like this helps to compare the distribution of scores between boys and girls. If just looking at the boxplots in Figure 5 then essentially the same picture emerges as found originally with the back-to-back histograms. Thus, while there is a slight difference (indicated by the slightly higher median line in the box for girls compared to boys), the two boxplots illustrate the fact that there remains a considerable overlap.

With this in mind, it is therefore now possible to illustrate the differences in GCSE Scores gained by the six different ethnic groups of pupils. As can be seen from Figure 6, ethnicity is clearly associated with levels of educational attainment as seen by the differing median scores for each of the six groups. As before, the differences are certainly sufficient for it to be a cause for concern. However, putting the boxplots side-by-side like this illustrates the fact that there is no discrete ‘underachieving cluster’ of Black, Pakistani and Bangladeshi pupils. While they certainly have the lowest median scores, there is so much overlap between them and the other groups in terms of the range of scores achieved that it makes little sense to describe these as a distinctive cluster and it is certainly misleading to continue to talk of a ‘gap’ between these and the rest at this level of analysis. Again, this is an important corrective to ecological fallacy that can sometimes arise from the interpretations of the summary statistics and results of the significance tests outlined earlier.

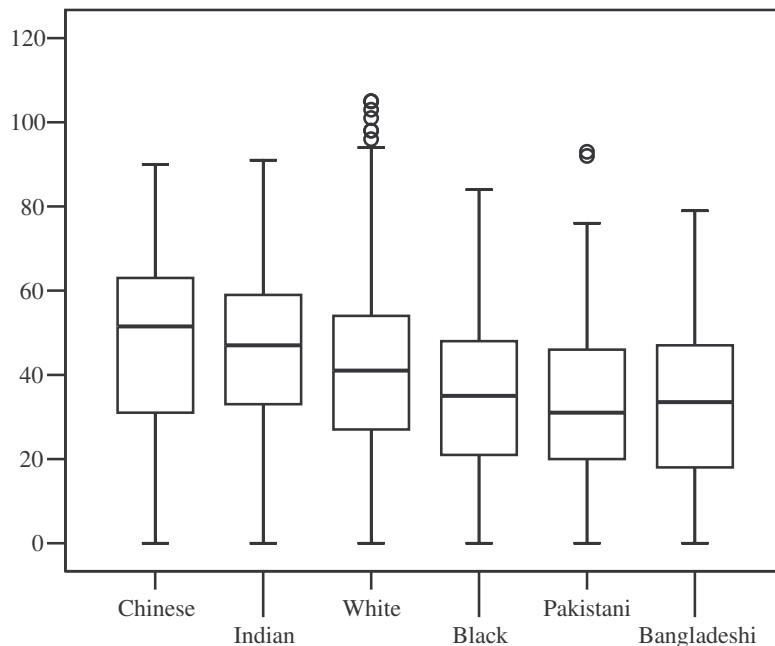


Figure 6. Boxplots Illustrating Ethnic Differences in GCSE Scores Among School Leavers in England and Wales in 1999

## CONCLUSIONS

What this paper has attempted to show are the potential problems and dangers associated with an over-reliance on the use of summary statistics and significance tests to describe and analyse inequalities in educational attainment. In showing this the paper is not suggesting that there is no need for such forms of analysis. Clearly, the use of summary statistics is essential if we are to identify and begin to understand structural inequalities and the broader processes that underpin them within education. Moreover, they play a vital role in monitoring progress over time. Rather, the key point here is simply that if they are used in isolation, which they often are, then they can encourage the tendency to apply group-level differences to individuals. The dangers associated with this have been clearly shown in relation to the political consequences of the moral panic over boys' underachievement as well as the tendency to label whole groups of minority ethnic pupils as underachievers.

It is with this in mind that the paper has argued that the use of summary statistics needs to be complimented with an emphasis on the graphical presentation and analysis of the data. While only a handful of the many techniques and methods associated with exploratory data analysis (EDA) have been used here by way of illustration, they have clearly demonstrated how such graphical displays can successfully illustrate the complexity of the data and the considerable variations within groups and overlap between them that exist in practice. In doing this, such methods are well placed to help challenge and counteract the ecological fallacy. It is also in this sense that quantitative data analysis does not necessarily have to be equated with a crude and simplistic approach to educational research. There is no reason why the complexity of the data cannot be drawn out and highlighted. As mentioned earlier, it is only because we tend to associate quantitative analysis solely with statistics that we have not as yet benefited from the insights to be gained through the use of EDA. Ironically, it is only when we begin to analyse quantitative data a little more qualitatively in this way will we be able to fully contextualise the summary statistics we produce and thus effectively counter the ecological fallacy once and for all.

## REFERENCES

- Arnot, M., David, M. & Weiner, G. (1999) *Closing the Gender Gap: Postwar Education and Social Change* (Cambridge, Polity Press).
- Bryman, A. & Cramer, D. (2001) *Quantitative Data Analysis with SPSS Release 10 for Windows* (London, Routledge).
- Connolly, P. (2004) *Boys and Schooling in the Early Years* (London, RoutledgeFalmer).
- Demack, S., Drew, D. & Grimsely, M. (2000) Minding the gap: ethnic, gender and social class differences in attainment at 16, 1988-95, *Race Ethnicity and Education*, 3, 117-143.
- DfES (Department for Education and Skills) (2002) *Statistics of Education: GCSE/GNVQ and GCE A/AS Level Performance of Candidates Attempting Two or More GCE A Levels of AS Equivalent in 2000/01*, Issue No. 01/02 (London, DfES).
- Epstein, D., Elwood, J., Hey, V. & Maw, J. (1998a) *Failing Boys? Issues in Gender and Achievement* (Buckingham, Open University Press).

- Epstein, D., Elwood, J., Hey, V. & Maw, J. (1998b) Schoolboy frictions: feminism and “failing” boys, in: D. Epstein, J. Elwood, V. Hey & J. Maw (1998) *Failing Boys? Issues in Gender and Achievement* (Buckingham, Open University Press).
- Gillborn, D. & Gipps, C. (1996) *Recent Research on the Achievements of Ethnic Minority Pupils* (London, Office for Standards in Education).
- Gillborn, D. & Mirza, H. (2000) *Educational Inequality – Mapping Race, Class and Gender: A Synthesis of Research Evidence* (London, Office for Standards in Education).
- Gorard, S. (2001) *Quantitative Methods in Educational Research: The Role of Numbers Made Easy* (London, Continuum).
- Gorard, S. (2003) Understanding probabilities and re-considering traditional research training, *Sociological Research Online*, 8, <<http://www.socresonline.org.uk/8/1/gorard.html>>.
- Hartwig, F. & Dearing, B. (1979) *Exploratory Data Analysis*, Sage University Paper Series on Quantitative Applications in the Social Sciences, 16 (Beverly Hills, Sage).
- Haque, Z. (2000) The ethnic minority 'underachieving' group? investigating the claims of 'underachievement' amongst Bangladeshi pupils in British secondary schools, *Race Ethnicity and Education*, 3, 145-68.
- Hoaglin, D., Mosteller, F. & Tukey, J. (eds) (1983) *Understanding Robust and Exploratory Data Analysis* (New York, John Wiley and Sons).
- Jackson, D. (1998) Breaking out of the binary trap: boys' underachievement, schooling and gender relations, in D. Epstein, J. Elwood, V. Hey & J. Maw (eds.) *Failing Boys? Issues in Gender and Achievement* (Buckingham, Open University Press).
- Lucey, H. & Walkerdine, V. (2000) Boys' underachievement: social class and changing masculinities, in: T. Cox (ed.) *Combating Educational Disadvantage: Meeting the Needs of Vulnerable Children* (London, Falmer Press).
- Mac an Ghail, M. (1996) ‘What about the boys?’: schooling, class and crisis masculinity, *The Sociological Review*, 44, 381-397.
- Rampton Report (1981) *West Indian Children in Our Schools* (London, HMSO).
- Raphael Reed, L. (1999) Troubling boys and disturbing discourses on masculinity and schooling: a feminist exploration of current debates and interventions concerning boys in school, *Gender and Education*, 11, 93-110.
- RSGB (Research Surveys of Great Britain) (2001) *Youth Cohort Study Cohort 10, Sweep 1 (C10S1): Technical Report* (London, RSGB).
- Troyna, B. (1984) Fact or artefact? The “educational underachievement” of black pupils, *British Journal of Sociology of Education*, 5, 153-166.
- Tukey, J. (1977) *Exploratory Data Analysis* (London, Addison-Wesley).
- Yates, L. (1997) Gender equity and the boys debate: what sort of challenge is it?, *British Journal of Sociology of Education*, 18, 337-47.