

Is discrimination inherent in the British teaching profession? **Stephen Gibbons** and **Arnaud Chevalier** investigate differences between the assessments that teachers give pupils and pupils' test marks to find out if teachers consistently underestimate the ability of pupils from certain backgrounds.

# Assessing pupils' abilities: do teachers and tests disagree?

**F**amily background remains one of the most powerful forces driving academic achievement and life chances in Britain. A large body of evidence shows systematic differences in achievement according to pupils' ethnic and socio-economic background. The reasons for these differences are not fully understood. But one undercurrent of opinion maintains that at least some of the differences could be due to failures or biases in the assessment system rather than any real differences in ability.

One concern is that teachers may, inadvertently or otherwise, stereotype pupils when making face-to-face assessments of their abilities. For example, they may judge individual pupils from Asian or black ethnic minorities based on preconceived notions of the average ability of Asian or black pupils. This is particularly worrying as there have been accusations of institutional racism in England's schools, particularly linked to exclusions of ethnic minorities.

There is a great deal of evidence that people engage in stereotyping or 'statistical discrimination' in all walks of life. But this should be of even more concern if it affects pupils from already disadvantaged ethnic or socio-economic groups. This might happen if, for example, teachers' assessments influence pupils' education and life trajectories through the number

and type of qualifications entered, through the feedback that teachers give pupils about their own abilities, and through academic references.

Our research looks at pupils participating in national tests at the ages of 11 and 14 in England. Because these pupils are assessed both by their teachers and by externally marked tests, we can compare the assessment that teachers give with the test marks that pupils receive in English, maths and science.

There is no reason to expect tests to

There is no evidence of institutional racism in the system of teachers' assessment of pupils



give a more accurate picture of ability than teachers' assessments or vice versa. But there is also no reason for the two methods to differ systematically according to the ethnic, gender or socio-economic group of the pupil being assessed. Evidence of this kind of divergence would suggest something is amiss in the assessment system, lending credence to the idea that stereotyping is pervasive in England's schools.

Our research is the first to look at this issue in the context of the full population of England's pupils at age 14. It is also the first to be able to take account fully of pupils' previous achievements, background characteristics, place of residence and school attended. And we do find that teachers' assessments and externally marked tests tend to diverge systematically according to the characteristics of the pupil being assessed.

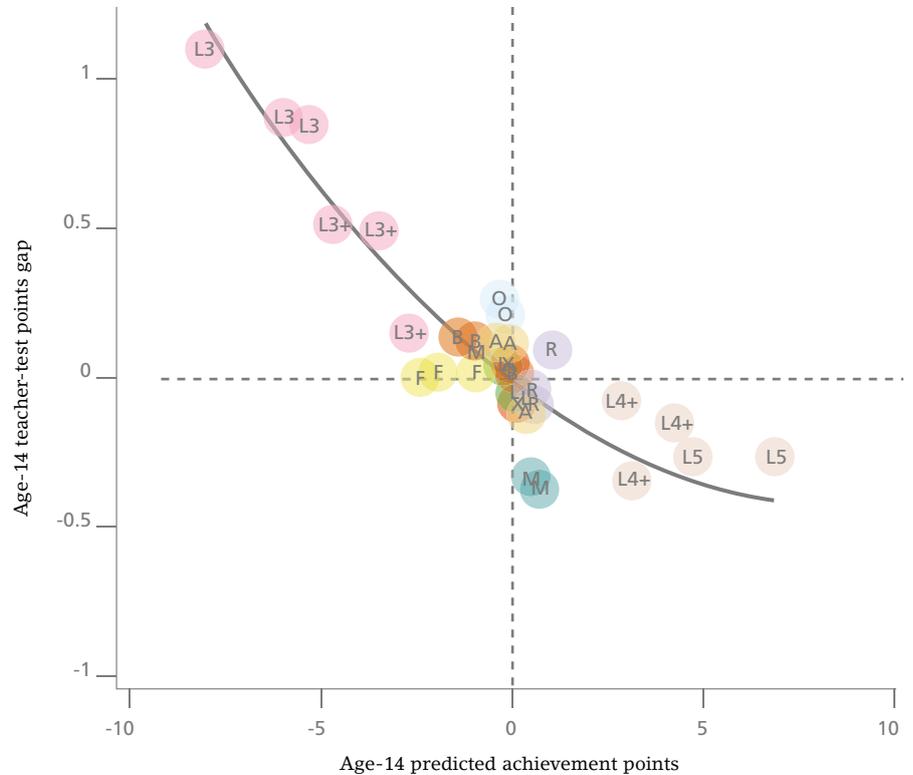
But this divergence does not happen in a way that is consistent with stories of statistical discrimination. If anything, teachers' assessments tend to work in favour of pupils who would be predicted to do relatively poorly on the basis of past assessments and the performance expectations of their demographic group.

We find, in other words, that higher ability pupils tend to be graded higher by the tests than by the teachers and low-achieving pupils better by the teachers' assessments than by the tests. Our data do not allow us to find out why this is the case. One likely explanation is that teachers (like most other people) have a tendency to extreme aversion in decision-making – that is, the tendency to go for intermediate rather than extreme decisions in the face of uncertainty.

Figure 1 demonstrates this finding: the horizontal axis plots pupils' predicted score in the tests they sit at age 14; the vertical axis plots the difference between teachers' assessments and test scores. All scores are scaled so that zero corresponds to an August-born white girl, not on free meals, with English as her first language, and who scored Level 4 on both teacher and test assessments at age 11.

Each data point has a label designating a pupil group. The L labels correspond to achievements at age 11 with L3 the lowest and L5 the highest. The other symbols are: F free meals, B black, A Asian, X mixed ethnicity, R other ethnicity, L English additional language, M male, O older (born in September)

Figure 1: The relationship between teacher-test points gaps and age-14 predicted achievement points by Key Stage 2 achievement level and different demographic, ethnic and free meal groups



- Key**
- F Free meals
  - B Black
  - A Asian
  - X Mixed ethnicity
  - R Other ethnicity
  - L English additional language
  - L3 L3+ Low test score stage 11
  - L4+ L5 Higher test score at age 11
  - M Male
  - O Older (born in September)



Higher ability pupils tend to be graded higher by the tests than by the teachers – and low-achieving pupils better by the teachers than by the tests

(birthday in September). There are three data labels of each type, corresponding to results in English, maths and science.

Data points in the top half of the figure represent pupil groups who do better in the teachers' assessments at age 14 than they do in tests. Data points in the bottom half represent pupil groups who do relatively well in the tests. Data points in the right hand side represent pupil groups who do better in both test and teacher scores. Data points on the left hand side represent lower achieving pupil groups.

The most striking feature of the figure is the obvious downward trend, with some very substantial gaps between teacher and test scores at age 14 with respect to predicted achievement. Pupils who scored towards the bottom of the distribution at age 11 (L3, L3+, top left quadrant) do relatively well on the teachers' assessments at age 14, while their peers at the top of the achievement distribution (L4+, L5, bottom right quadrant) do relatively well in the tests. The differences by free meal entitlement, ethnic group and demographics are modest in comparison although they follow the same general trend.

Should we worry about these gaps even though they do not seem to correspond to traditional views of stereotyping? And could these gaps between teachers' and test assessments have any bearing on what happens to

children in the future in terms of their academic success?

To answer these questions, we examine whether pupils who score well on teachers' assessments relative to tests are entered for more GCSEs, do better in their GCSEs, choose different GCSE subjects or are more likely to stay in education after the age of 16. In no case could we find any convincing evidence that discrepancies between teachers' and test assessment scores had any meaningful influence on any of these outcomes.

We cannot say from this research whether there are emotional effects from bad test results or teacher assessments. But evidently, pupils' academic performance does not seem to suffer in the medium term as a consequence of assessment biases or errors.

In another strand of related research, we find that pupils who are under-confident in their abilities are less likely to expect to go to university. But university graduates' lack of confidence about their scores in specific cognitive tests has little connection with how well they think they will do in their exams or their expected success in the labour market. The two studies together suggest that while academic ability matters a lot for subsequent outcomes, personal judgements about ability – whether by teachers or students themselves – do not always have a big role to play.

Although there is no evidence of

## Errors in school assessments have little influence on pupils' GCSEs or decisions about education after age 16.

institutional racism in the assessment system or that standard forms of stereotyping by teachers is going on, there are nevertheless systematic differences in the way tests and teachers rate pupils of high and low abilities. This raises some questions about the overall reliability of the assessment system in England as it stands. But more importantly, the systematic discrepancies suggest it would be very unwise to move to a system that was totally reliant on one form of assessment alone, either teacher or test-based.

Even so, the research results suggest that it is unlikely that pupils' long-term school performance is heavily influenced by teachers' perceptions of their abilities or by any other form of bias in school assessment. Nor do these factors seem to be a big influence on pupils' decisions about staying in school after 16 or gaining the prerequisite qualifications for participation in higher education.

This article summarises 'Assessment and Age 16+ Education Participation' by Stephen Gibbons and Arnaud Chevalier, *Research Papers in Education* 23(2): 113-23 (June 2008). The research was carried out for the Widening Participation in Higher Education project funded by the ESRC's Teaching and Learning Research Programme.

The study of university graduates is 'Students' Academic Self-perception' by Arnaud Chevalier, Stephen Gibbons, Sherria Hoskins, Martin Snell and Andy Thorpe, CEE Discussion Paper No. 90 (<http://cee.lse.ac.uk/cee%20dps/ceedp90.pdf>).

**Stephen Gibbons** is a senior lecturer in economic geography at LSE.

**Arnaud Chevalier** is a senior lecturer in economics at Royal Holloway, University of London. Both are research associates in CEP's education and skills programme.

