

ANALYSING EXAMINATION

RESULTS

CAROL TAYLOR FITZ-GIBBON

1990

8 Analysing Examination Results

C. T. FITZ-GIBBON

At the time of writing, LEAs generally present examination results as 'raw data', perhaps with some indication of factors which might be taken into account in their interpretation. Sometimes the data are compared with available averages, such as the county average, or the national average. This approach has the comfort of vagueness and it might even be best if this kind of data were all that was required for public disclosure. Sometimes, however, despite the vagueness of the data, institutions are unfairly criticised. The problem which arises is: how is anyone to know to what extent a set of results 'ought' to be close to any particular average? In a vague sense it might be known that allowances for 'low' examination results should be made because the area in which a school is located is deprived, but the comparison is only vague, not quantified. Furthermore there are genuine problems in 'making allowances' because an area is 'deprived', as Chris Tipple points out with humour in Chapter 4. There are bright pupils in deprived areas and should not as much be expected of them as bright pupils in wealthy areas? This is an on-going issue on both sides of the Atlantic.

A more quantified approach to the interpretation of examination results rests on having some measure of developed academic abilities and then seeing if pupils are obtaining the results appropriate to these academic abilities. Measures of developed abilities or prior achievement are, in any case, better predictors of examination results than measures of home background or socio-economic status.

An HMI report on the use that LEAs make of examination results (described in the Annotated Bibliography) confirms that very few LEAs analyse examination results in this more sophisticated way, making

allowance for developed abilities. It is, however, the approach adopted by researchers seeking to estimate the effects that schools have on pupils.

Perhaps it should be emphasised that we are not advocating that the 'abilities' of pupils should be known to teachers on an individual basis. The tests are obviously not 100% accurate and have clear limitations as to what they measure. Nevertheless, as a way to make fairer comparisons among *groups*, the tests are useful. Some adjustments made to examination results for levels of ability is desirable in order to be fair to the schools involved.

It is not difficult to make allowances, statistically, for developed abilities, so that schools are compared on the basis of how good their examination results are given the abilities of the pupils they are working with. How these allowances are made by educational researchers will be described shortly, but first let us consider why the ubiquitous percentage pass rate is one measure of examination results that researchers generally do *not* use.

Defects of Percentage Pass Rates as Performance Indicators

The defects of percentage pass rates are spelt out below with the intention of showing why they should not be used to judge schools and colleges.

No account is taken of the kinds of pupils entered

This lack of adjustment for intake is the most serious problem and applies to any raw set of examination results, such as numbers of passes at various levels. Schools should not be judged on raw results which directly compare, for example, schools which have pupils of higher ability with schools which, for whatever reason, have pupils of lower abilities. A 60% pass rate might sometimes represent a better achievement on the part of the school than a 90% pass rate. It all depends on what could reasonably be expected.

Barely passing is counted in the same way as getting a high grade

When percentage pass rates are used, the focus is purely on the pass/fail dichotomy. Since we know that the level of a pass is important to people who

use the results (e.g. employers and selectors for higher education) the reliance on a simple dichotomy is unjustifiable. University and polytechnic admissions tutors, for example, look closely at the levels of A-level passes and indeed add up results on the UCCA scale, counting 5 for an A, 4 for a B and so on.

A booklet published by the Chartered Institute of Public Finance and Accountancy (CIPFA, 1988) suggests both 'percentage of age cohort obtaining 5 or more results at grade F or better in GCSE' and 'percentage of age cohort obtaining 4 or more results at grade C or better in GCSE', thus making some distinction between different levels of pass but in so doing another peculiar and confusing error is introduced: double counting. The second figure is partially subsumed in the first figure.

Adding up passes in all subjects loses information

If passes are added up from all subjects information about the different departments in the schools is lost. School departments may differ in effectiveness and to ignore this variation is to lose important information for planning improvements.

A distorting effect on educational practice may occur

An 'indicator' suggested by CIPFA is 'Percentage of pass rate (A-E) at A-level' (*sic*). This illustrates another problem with percentage pass rates: their use may push institutions to allow only potentially successful candidates to attempt some examinations such as A-levels. By only taking on students who seem sure to pass, the pass rate can be kept high. In what are to a large extent norm-referenced examinations this tactic could not succeed if adopted by everyone and is therefore a tactic to be discouraged. At A-levels for example, about 30% are predictably going to 'fail' so it is reasonable to expect most institutions with average intakes to take on a sacrificial 30%. If percentage pass rates are used, then A-level institutions with higher than a 70% pass rate should be asked if they are pulling their weight in this respect, not praised! (Of course, there is much to question in a system which 'fails' 30% of its academically most able students but that is a another problem.)

The percentage pass rate ignores the curriculum balance

Not all examinations are equally difficult. For example there is evidence to suggest that a D in A-level mathematics is about as academically difficult to get as a B in A-level English. A school or college which, by attracting many candidates into mathematics, helps to ameliorate the national problem of a shortage of numerate persons might thus find itself at a disadvantage if percentage pass rates were examined.

Percentage of what?

An inevitable question is 'percentage of what?' In the examples given from CIPFA it can be seen that the answer is sometimes but not always 'percentage of the age cohort'. But there is room for vagueness on this quantity, a vagueness which allows the indicator to be strategically manipulated. When simple percentage pass rates are called for this presumably implies the percentage is the percent of those entered. Schools and colleges are vague about this quantity. What of pupils who fail to show up for the examination? Are adults counted? What of those who take the examination through some other institution? Are resits included? Various problems with pass rates are encountered in the diverse provision currently available.

In summary, if an indicator like percentage pass rate is used institutions may be tempted to be guided in their advice to students by considerations relating to the performance indicator rather than relating to the students' long-term prospects, educational needs and interests. Clearly we need better indicators for examination performance, indicators which do not unduly influence what should be educational decisions and indicators which are arrived at subject by subject so that they are not influenced by curriculum choices. We need also indicators which value each student equally because we want teachers to act with that fairness of approach. Double counting higher passes implies an unfair concentration on some students more than others.

The only virtues of percentage pass rates are that they appear simple. But simplicity is not the prime concern for performance indicators. Good indicators are fair and beneficial, promoting good educational decision making.

Example of a Fairer and More Beneficial PI

In assessing examination results each child should 'count' equally, both in fairness to the schools being monitored and in fairness to the child. How is a fairer indicator for examination results to be calculated? To compute fair indicators some measure of the pupils is required as an 'intake' measure. This might be, for example a measure of developed abilities or prior achievement. It must be a measure which is fairly strongly correlated with examination results.

When we have such an intake measure we can compare the examination results obtained by any group of students with those of other students who had the same intake characteristics. In other words, given an intake measure which 'predicts' examination results, then we know roughly what grades to expect for various groups of students. If many students in one school achieved better than would have been expected of them given the general pattern of results, we might consider this finding to be a positive indicator for the school. For example, even though all students got Es this could be a very good result if, in other schools, pupils of similar abilities had all failed.

How can we know what to expect of a student? Obviously we cannot know accurately for one student but if we consider, say, a few hundred students, we can make some predictions based on the general pattern. To invent a simple example, suppose over several hundred students we found that the pattern relating GCSE-grade to A-level grade was this: at A-level a student tends to get one grade lower than at GCSE. We would then have a 'prediction equation':

$$\text{A-level-grade} = \text{GCSE-grade minus one}$$

OR

$$\text{Agrade} = \text{GCSEgrade} - 1$$

The outcome (A-level grade: 'Agrade') is predicted from the intake measure ('GCSEgrade'). The relationship between GCSE grades and A-level grades is not quite so simple but it can be represented by an equation like the one above.

The essence of the idea is that if we know something about a candidate, such as his or her prior achievement, and if we know how the prior achievement in general, in all the participating schools and colleges, related to A-level grades, then we have the basis for predicting the A-level grade each candidate might have been expected to get. If the candidate got a grade higher than the one predicted then that is a positive point for the school. If the candidate got lower than predicted, that is a negative point.

The difference between the actual grade a student gets and the grade predicted for the student is called a residual. Positive residuals indicate better than predicted performance. Negative residuals represent worse than predicted performance. By averaging the residuals for a school we can see if, on average, a school's results are better or worse than would have been predicted on the basis of the prior achievement of its candidates. The averaging should be done subject by subject.

Of course, a particular candidate's over- or under-achievement may be nothing to do with the school but if a pattern builds up of, say, most candidates doing better than expected, then it would look like a positive indication. It is certainly a fairer indication than percentage of passes: it uses all the information (not just a pass/fail dichotomy) and enables outcomes to be compared having taken account of prior achievement. Statistically this process is referred to as 'controlling for' prior achievement.

The example of GCSE grades has been used for simplicity but of course other characteristics of the students could be 'controlled for': home background, ability, other measures of prior achievement. Indeed, several predictors can be 'controlled for' (taken into account) at once. The procedures for doing this are known in statistics as multiple regression analysis. Many statistical packages can produce such analyses and as schools move into administration based on computer networks there is every reason to produce performance data on examination results which are at least as fair as this and to move away from such problematic measures as percentage pass rates.

References

Chartered Institute of Public Finance and Accountancy 1988, *Performance Indicators in Schools*. London: CIPFA.