

DRIVING TESTS: RELIABILITY, AND THE RELATIONSHIP BETWEEN TEST ERRORS AND ACCIDENTS

Chris Baughan and Barry Sexton
Transport Research Laboratory
Crowthorne, Berkshire RG45 6AU
U.K.
E-mail: cbaughan@trl.co.uk

Summary: In the British practical driving test, serious or dangerous faults are those judged to involve potential or actual danger, and a single such fault results in test failure. As part of a wider project to review the driving test, TRL conducted a study of test-retest reliability. Test and retest outcomes differed for a substantial proportion of candidates. The paper argues that inconsistent performance on the part of the candidate is likely to explain much of this apparent unreliability.

Faults that are less serious are also recorded during the test but few candidates make the 16 of them necessary to fail. Self-reported accidents during the first six months of post-test driving were analysed together with driving test records for approximately 30,000 drivers. Statistical modelling suggested that people who pass the driving test having made large numbers of the less serious faults may be intrinsically less safe as drivers, but that they also tend to drive less overall, and less often at night. This reduces (and for men removes) the association between test faults and the actual number of accidents reported.

INTRODUCTION

In the British practical driving test, the examiner chooses one of several set routes. Driving is monitored for 48 categories of predefined faults, and examiners are trained to assess each as a dangerous, serious or 'driving' (i.e. less serious) fault. Candidates making one or more serious or dangerous faults, or more than 15 driving faults, fail the test. This paper describes results from two studies commissioned by the UK Department of the Environment, Transport and the Regions as part of a wider review of the British practical driving test. They are a study of test-retest reliability, and a study to investigate the relation between the number of driving faults recorded during the test, and accidents in the first six months of post-test driving.

THE RELIABILITY OF DRIVING TESTS

At first sight, the mechanism by which the test seeks to achieve a safe and competent driver population is that of driver selection or screening: drivers who do not meet the test standard are not allowed to drive unsupervised. In practice, only a small proportion of the population is permanently excluded from solo driving. Most people who fail the test simply take more training and practice, and retake the test; repeating the process until they pass. The test's main function is to influence the amount and quality of training and practice accumulated by learner drivers before they are allowed to drive solo. The importance of this function is underlined by

the fact that, in Britain, the licensing system does not require learner drivers to take any specified amounts of training or practice.

A test with zero reliability for actual candidates (i.e. people who book a driving test in the course of learning to drive) could tell us nothing about their competence or accident liability, and would provide no basis for withholding or granting a licence. However, the test could still have good ‘consequential’ validity (in the sense of having the desired beneficial effect on road safety and driver competence) if it induced candidates to have sufficient training and practice. To do this, the test would have to be reasonably reliable in failing *potential* candidates who are at an earlier stage of learning and who would be unsafe or incompetent as unsupervised drivers. If we accept the underlying error rate as an index of a candidate’s driving performance, the test would have to be reliable in failing potential candidates with error rates substantially higher than those typical for actual candidates.

Although a test with low reliability for actual candidates could be fulfilling its main function of maintaining standards of training, safety, and competence, high reliability is still desirable for reasons of efficiency and perceived fairness.

A Test - Retest Study

Candidates who had already booked a driving test volunteered to take a second driving test within a few days of the first. The 366 participants were not told the result of the first test, or given any feedback on how they had done, until after they had completed the second. The second examiner was not told the result of the first test until after the second test had finished. Table 1 shows the main results. Sixty four per cent of candidates received the same result in each test, 20% failed the first test and passed the second, and 16% passed the first but failed the second

Table 1. Results of test-retest study
Frequencies (% in brackets)

Result of 1st test	Result of 2nd test		Total
	Pass	Fail	
Pass	80 (22)	57 (16)	137 (37)
Fail	75 (20)	154 (42)	229 (63)
Total	155 (42)	211 (58)	366 (100)

Explaining Test-Retest Inconsistency

Between-examiner effects are unlikely to explain much of the inconsistency. Results of check tests with two examiners in the same car show very high levels of agreement on test outcome. Much of the test-retest inconsistency is almost certainly to do with candidates coming forward for test when they are only just good enough to pass if they drive at their best. In other words, their rates of making serious and dangerous faults are such that, though they may get away without making one during a given test, there is a reasonably high probability that they will make one in another test.

A Simplified Model of a Driving Test

Some insight into the effect of individual candidates' underlying competence on the reliability of a driving test can be gained by examining a simplified model.

In this model, the level of driving competence reached by a test candidate is characterised by that individual's underlying expected error rate. For the purposes of the model, a candidate's errors are assumed to be generated by a Poisson process – i.e. a candidate's expected error rate is the mean of a Poisson distribution. Such a model may be used to calculate maximum test-retest agreements (i.e. assuming no other sources of unreliability) for candidates with given expected error rates. Table 2 shows the results of such calculations, and indicates that test-retest agreements remain rather low over a wide range of expected error rates.

Table 2. Percentage of test-retest agreements for different expected error rates

Failure criterion	Expected errors per test (Poisson intensities)							
	0.1	0.2	0.5	1	2	3	4	8
1 error	83%	73%	52%	53%	77%	90%	96%	100
2 errors	99%	96%	84%	61%	52%	68%	83%	99%

Improving the Reliability of the Driving Test

In principle, this could be addressed by increasing the numbers of errors observed during the test, i.e. by increasing the duration of the test or by basing the test on errors that are less severe and therefore more frequent. Such a change was made to the test in May 1999, based on evidence that minor errors show some ability to detect candidates who have a raised accident liability (Maycock and Forsyth, 1997). On the above argument, it could also improve reliability. Initially the failure criterion was set at the high value of 16 driving faults – which has meant that less than half of one percent of candidates fail on this basis. The intention was to tighten this criterion if it could be shown that to do so would fail people with a raised accident liability.

TEST FAULTS AND ACCIDENT LIABILITY

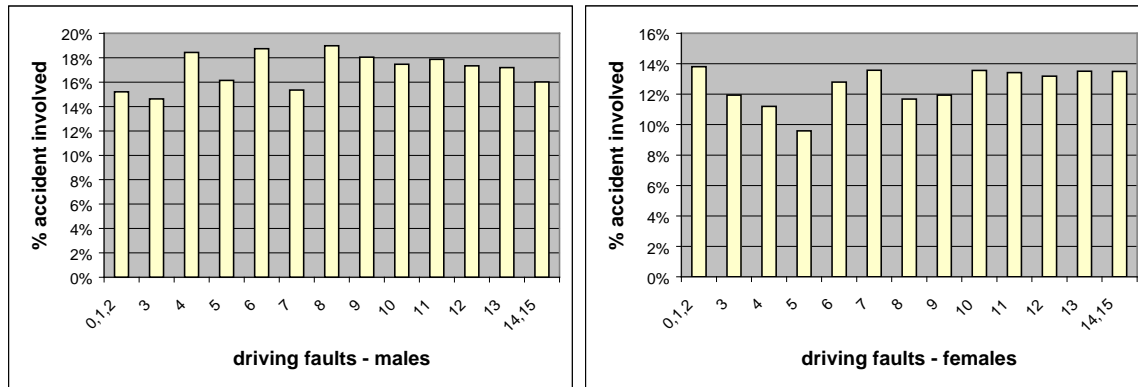
Drivers who had passed the test about seven months earlier were asked to complete a postal questionnaire, covering accidents during the first six months of post-test driving and a number of other questions relating to exposure, driving style, driving competence, attitudes towards driving violations, and driving behaviours. 29,559 matched driving test and questionnaire records were available for the main analyses.

Figure 1 shows the percentages of men and women in each number-of-faults group who reported having been involved in an accident in the first six months of driving after passing the test. There was no obvious pattern of accident involvement being associated with high numbers of driving faults, though in fact it is possible to divide the women into higher and lower faults groups, with the higher faults group having a statistically significantly higher accident involvement.

Fitting Accident Models

People who made more driving faults tended to be a little older than the rest, and to drive rarely or never in the dark. This was true for both men and women, and suggested that high faults people might be intrinsically less safe as drivers, but might have no more accidents than low faults

Figure 1. Accident involvement and driving faults



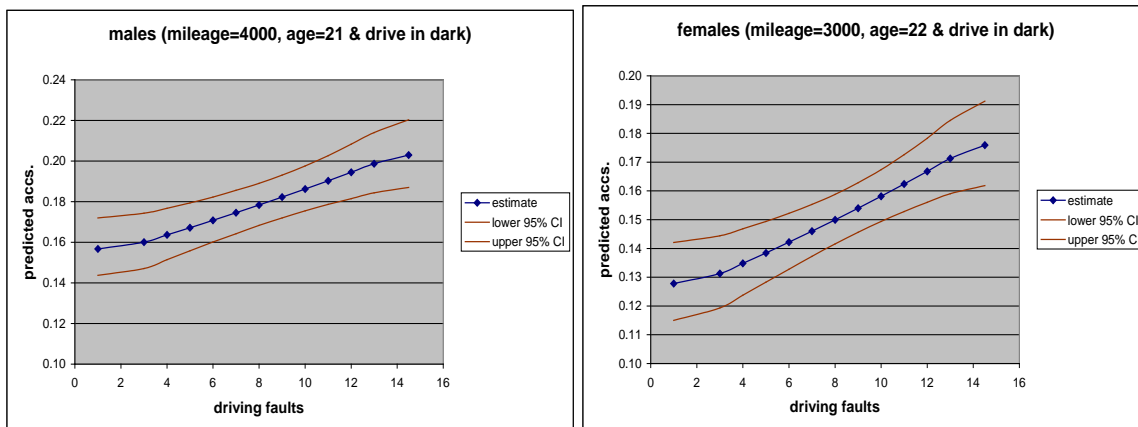
people because they have less exposure to risk. Generalised linear models of the following functional form were fitted to the data:

$$\text{Accidents} = \text{miles}^{b_1} * \exp\{ b_2/\text{age} + b_3*(\text{drive in dark}) + b_4*(\text{driving faults}) \}$$

The driving faults term in the model was statistically significant, indicating that there is a relation between faults and accidents once the effects of age, mileage and driving in the dark are adjusted for. Figures 2a and 2b illustrate this relationship for male and female drivers.

Figure 2. Predicted number of accidents in the first six months

- (a) for males aged 21 who drive 4000 miles, and drive 'often or sometimes' in the dark,
- (b) for females aged 22 who drive 3000 miles, and drive 'often or sometimes' in the dark



The modelling results are consistent with the argument that high-faults people are intrinsically less safe as drivers, implying that they would have more accidents than low-faults people if they

were to decide to drive the same mileages, and as often at night, as low faults people. The results imply that there is a case for failing high-faults drivers on the grounds that their driving is relatively unsafe and might be improved by further training and practice. Allowing them to pass the test is to rely, it could be argued, on a control mechanism (i.e. the self-controlling of exposure) that is not understood, and that could presumably be overridden at any time by a driver who so chooses.

Whether this argument is sufficient to justify failing high-faults people might be challenged for two reasons. First, it is clear from the study that high-faults people tend to limit their driving (or to have it limited by factors outside their control) such that they do not actually have more accidents than low-faults people. Secondly, exposure and driving experience are highly correlated in this sample of novice drivers, with mileage being the main indicator of both. In this cross-sectional sample the net effects of mileage and night driving were found to act as exposure variables rather than experience variables, but we cannot be sure that this would be so for individual drivers who decide to drive more.

CONCLUSIONS

Test-retest reliability of the practical driving test is rather low, with inconsistent performance on the part of the candidate being an important source of test-retest disagreements. This does not mean that the test is failing to induce good training and practice, or that there is anything wrong with the way the test is carried out. Nevertheless, an improvement in reliability is desirable. Encouraging learners to have more driving experience before coming for test should be helpful here.

It appears that high-faults people tend to be intrinsically less safe as drivers during the first six months of post-test driving, implying (though not providing absolute proof) that they would have more accidents than low-faults people if they were to decide to drive the same mileage and to drive as often at night.

High-faults people tend to drive fewer miles in the first six months after the test, drive less often the dark, and be slightly older than low-faults people. These differences tend to reduce the actual accident involvement of high-faults people. There is no relation between number of faults and actual accidents for men, and only a rather weak one for women.

ACKNOWLEDGEMENTS

The work reported here was sponsored by the Department of the Environment, Transport and the Regions, and carried out in TRL's Safety Group. The authors gratefully acknowledge the contribution made by TRL colleagues, and the invaluable help provided by Robin Cummins (Chief Driving Examiner), Trevor Wedge, and other Driving Standards Agency staff.
Copyright TRL Limited 2001

REFERENCE

Maycock, G. and Forsyth, E. (1997). *Cohort study of learner and novice drivers – Part 4: Novice driver accidents in relation to methods of learning to drive, performance in the driving test and self-assessed driving ability and behaviour.* TRL Report 275. Transport Research Laboratory, Crowthorne.