Novice driver safety and the British practical driving test

Prepared for Road Safety Strategy Division, Department for Transport

C J Baughan, B Sexton, G Maycock, H Simpson, L Chinn and A Quimby
CONTENTS

Executive Summary 1

1 Introduction 9

2 Assessing driving tests 9

2.1 The British driver licensing system 9
2.2 Objectives and functions of the driving test 9
2.3 Validity, reliability and pass-rates 9
2.3.1 Reliability and validity 9
2.3.2 Pass-rate 11
2.3.3 Causes of low pass-rates 11
2.4 Conclusions on assessing driving tests 11

3 The British practical driving test: why consider changing it? 12

3.1 The practical driving test 12
3.2 Indications of a possible need for change 13
3.2.1 Modern traffic conditions 13
3.2.2 Novice driver accident liability 13
3.2.3 Novice driver accident types 13
3.2.4 Research into driver behaviour 13
3.2.5 Opinions of examiners, instructors and candidates 14
3.2.6 Test marking systems 14
3.2.7 Other considerations 14
3.2.8 European Commission requirements 14

3.3 Summary of types of change considered for the practical driving test 15
3.3.1 Make better use of less serious faults as a failure criterion 15
3.3.2 Add to the range of driving conditions covered 15
3.3.3 Increase the duration of the test 15
3.3.4 Improve the use made of the test as a training opportunity 15
3.3.5 Improve the standard that candidates choose to reach before coming for test 16
3.3.6 Improve coverage of higher order skills such as hazard perception 16
3.3.7 Examiners’ ratings of performance 16
3.3.8 Other fundamental changes to the marking system 16
3.3.9 Add attitudinal items that predict future unsupervised driving behaviour and accidents 16
3.3.10 Test the manoeuvres off-road 16
3.3.11 Increase the pool from which ‘called for’ manoeuvres may be selected 16
4 Opinions of examiners and instructors on the test and how it might be improved

4.1 Exploratory interviews with DSA driving examiners

4.1.1 Possible improvements to the test

4.1.2 Other issues

4.1.3 Conclusions from the exploratory interviews

4.2 Group discussions with examiners and instructors

4.2.1 Possible improvements to the test

4.2.2 Other issues

4.2.3 Conclusions from examiner and instructor group-discussions

5 Factors influencing the pass-rate of the practical driving test

5.1 Introduction

5.2 Data collection

5.2.1 Sample of test centres

5.2.2 Examiner questionnaire

5.2.3 Route characteristics

5.2.4 The combined data

5.3 Results

5.3.1 Candidate age and sex, and time-of-test effects

5.3.2 Examiner, centre and regional effects

5.3.3 The effect of individual route characteristics and driving conditions on pass-rates

5.4 Discussion

5.5 Conclusions on factors influencing pass-rates

6 Test reliability and consistency of candidates’ performance: a test–retest study

6.1 Introduction

6.2 Study design

6.3 Results

6.4 Explaining test–retest inconsistency

6.5 Improving the reliability of the driving test

6.6 Conclusions on test reliability

7 How L-Drivers decide when to take the practical driving test – an exploratory study

7.1 Introduction

7.2 Method

7.3 Results

7.4 Discussion

7.5 Conclusions on candidates’ decisions about when to take the test
8 Novice driver accidents

8.1 Introduction 42
8.2 Summary of findings 43
8.3 Implications for the practical driving test 43

9 Faults made during the test, and accidents in the first six months of driving

9.1 Introduction 45
9.2 The survey 46
  9.2.1 Sample design 46
  9.2.2 The questionnaire 46
  9.2.3 Fieldwork 46
9.3 Results 46
  9.3.1 Achieved sample 46
  9.3.2 Proportions of drivers involved in accidents 46
  9.3.3 Can a driving faults threshold identify a group who go on to have more accidents than other drivers? 46
  9.3.4 Age and exposure of high-faults drivers 48
  9.3.5 Fitting accident models 48
  9.3.6 Practical significance of the modelling results 48
  9.3.7 Responsibility for accidents 49
  9.3.8 Low speed manoeuvring accidents 49
  9.3.9 Accidents and types of driving fault 49
  9.3.10 Factor and discriminant analyses of driving faults 50
  9.3.11 Characteristics of high-faults drivers 50
9.4 Discussion 51
  9.4.1 Does the survey support changing the failure criterion in the driving test? 51
  9.4.2 Targeted vs. general safety interventions 51
9.5 Conclusions on tightening the failure criterion for driving faults 52

10 Other ways of using the licensing system to improve safety 52

10.1 Introduction 52
10.2 Components of graduated licensing 53
  10.2.1 Extending the pre-solo learning phase 53
  10.2.2 Night driving restrictions 55
  10.2.3 Informal supervision by parents or other adults 56
  10.2.4 Passenger bans/restrictions 57
  10.2.5 Reducing alcohol limits 58
  10.2.6 Increased consequences of traffic violations 58
  10.2.7 Increasing quality or quantity of education/training 58
  10.2.8 Restrictions on speed or on using high speed roads 59
Executive Summary

Introduction
Research has shown that the accident liability of novice drivers decreases very sharply during the first six months of post-test driving, and continues to decline markedly over the next 2 – 3 years. Much of this is known to be associated with the gaining of driving experience rather than age (eg Maycock et al., 1991; Maycock 2002). This implies that safety could be improved by (a) enhancing the learning process; (b) preventing people from driving unsupervised until it has taken effect and (c) influencing behaviour or reducing exposure to risk during the early months of solo driving to counter the effects of inexperience and immaturity.

One approach here is to modify the driving test to improve the training and experience accumulated by learner drivers, and screen-out drivers who have not yet reached a standard acceptable for unsupervised driving. The Department for Transport commissioned TRL to assess the scope for such improvements to the practical driving test for car drivers.

An early task was to assess the evidence for and against a number of possible changes to the driving test, to see whether any of them could be recommended for adoption without further research. This work identified changes likely to improve the test and, after a public consultation, the Driving Standards Agency introduced some of them in May 1999. The project was then asked to conduct research on further aspects of the driving test, with a view to making recommendations for longer term improvements.

Studies were undertaken of the consistency of candidates’ performance at the time they come for test, the way in which candidates decide when to come for test, the literature on novice driver accidents and its implications for driver testing, the influence of route and other characteristics on the pass rate, and the relation between the less serious faults made by people who pass the test and their subsequent accidents.

In Britain, the driving test is the main tool for inducing learner drivers to build up training and experience before driving unsupervised. This is asking a lot of a driving test; it is also difficult for a test to include in its pass/fail criterion those variables that govern the discrepancy between supervised driving performance during the test and the subsequent unsupervised driving behaviour of novice drivers. It is therefore necessary to consider whether other changes to the training/testing/licensing system would be beneficial. Graduated licensing systems used in other countries aim to provide a staged progression from initial learning to unrestricted solo driving. They include measures designed to restrict exposure during early driving, exert a supervisory influence over driver behaviour during the first part of a driver’s solo driving career, or improve the level of training and experience accumulated before driving solo without restrictions. The project included a review of graduated licensing and related systems in other countries with a view to identifying measures that might be effective in Britain.

This report summarises some of the project subtasks is published elsewhere – (eg Maycock, 2002; Baughan and Simpson, 2002).

The British practical driving test before and after May 1999
Before May 1999, the British practical driving test lasted approximately 35 minutes. The examiner chose one of several set routes. The candidate had to execute an emergency stop, a hill start, and two out of three manoeuvres involving reversing (ie reverse parking, reversing into a limited opening, and a turn in the road). Driving was continuously monitored for 48 categories of predefined errors, each of which was assessed as a minor, serious, or dangerous fault. Serious or dangerous faults were those judged to involve potential or actual danger, and candidates making one or more such faults failed the test. There was no specified limit on the number of less serious faults a candidate was allowed to make without failing, although examiners could decide that a pattern of repeated less serious faults indicated a serious fault, resulting in failure. Only the first two of these less serious faults in any category were recorded by the examiner. There was no motorway driving, and many tests did not include driving on dual carriageways. Candidates at the end of the test were given a copy of the examiner’s marking sheet, and were offered a short verbal debrief. Instructors were encouraged to attend the debrief (and to ride in the back of the car during the test), but rarely did so.

The changes introduced in May 1999 were as follows:

- Test duration was increased by about seven minutes.
- Coverage of test routes was improved – more routes now included sections of higher speed dual and single carriageway roads.
- Changes2 to the testing of the emergency stop and reversing manoeuvres yielded three to five minutes more time for ‘free’ driving.
- All less serious faults (renamed driving faults) were now recorded.
- Candidates making more than 15 of these less serious faults now failed the test.
- Candidates were now given a written explanation of the test report at the end of the test, in addition to the marking sheet and verbal debriefing.

---

1 Department for Transport (DfT) is used throughout the report to include also the precursor Departments (Department for Local Government, Transport and the Regions, and Department of the Environment, Transport and the Regions).

2 The changes involved one in three candidates (rather than all candidates) being asked to demonstrate an emergency stop. Also, the reverse parking manoeuvre was now tested the test centre car park where facilities were available.
The British practical driving test: summary of current status and possibilities for future development

Overall effectiveness

Given the relative levels of road safety in Britain and elsewhere, it appears that the driving test has proved rather effective – ie it has helped to achieve a reasonably competent and comparatively safe driving population. However, the fact that novice drivers’ accident liability is high immediately after the test but falls sharply as they gain post-test experience shows that they still have a lot to learn after passing the test. If it is possible for such learning to take place before drivers come for test – and evidence from other countries indicates that it is – there would appear to be scope to improve further the effectiveness of the test. That is, it would be beneficial if the testing system induced more of this learning to take place pre-test than it does at present.

There are four main ways of modifying a driving test to improve the competence and safety of novice drivers. These are: (i) broadening the test content to induce candidates to achieve competence in a wider range of driving tasks; (ii) improving the performance of the test as a measuring instrument, so that it is better able to distinguish between candidates who have, and have not, reached a desired level of competence; (iii) increasing the severity of the test, so as to improve the level of competence achieved; (iv) making more use of the test itself as a training opportunity (ie via examiner feedback). Though conceptually distinct, in practice these categories overlap since a change that addresses one of them may affect others.

The status of the British practical driving test and the scope for improvement are discussed below, as is the question of whether changing the test alone would be the best way of achieving the desired benefits.

Test content

The test provides reasonable coverage of most of the basic driving situations and skills, and most test centres and many routes now have access to higher speed dual carriageways and national speed limit single carriageway roads. As a result the majority of learners will do some preparation on such roads even if their particular test does not include them. The accident data suggest that novice drivers have particular difficulties at night, and with judgement of appropriate speed (eg on the approach to bends and roundabouts in light traffic conditions). Ideally, therefore, the content of the test would be improved in these regards. However, speed selection on test is unlikely to be representative of the speed choices of unsupervised drivers. The difficulty of dealing with night driving in the driving test is obvious – and anyway, for novice drivers, night driving accidents may well be as much to do with factors associated with the recreational journeys being made as to problems of driving in darkness. It is therefore likely that the best solution lies in the training and licensing system, rather than within the test itself.

Novice drivers have relatively poor hazard perception skills. This situation should be improved by the recent introduction of a computer-based hazard perception test as part of the Theory Test that learner drivers have to pass before booking a practical test (Grayson and Sexton, 2002). The practical test also assesses such skills: people fail if poor hazard perception or poor observational procedures lead them to make a serious or dangerous fault. It might be possible to strengthen further the assessment of hazard perception skills in the practical test. Commentary drives, or questioning the driver about a situation just experienced, would be worth exploring, though validity and practical feasibility would need to be assessed. Strengthening this aspect of the practical test should encourage continued emphasis on hazard perception skills during practical training, and would require learners and their instructors to confront the difficulties of learning and using such skills at a time when basic control skills are occupying much of their attention.

There is growing realisation of the importance of drivers’ attitudes and goals as influences on behaviour, and the need for drivers to be aware of these influences and how to recognise and control them has implications for both training and testing. The practical test does not attempt to assess directly the attitudinal or motivational factors associated with the propensity of drivers to depart from good, safe driving practice once they begin to drive unaccompanied. However, it does not seem feasible to include such items in the pass/fail criterion because they are so open to manipulation by candidates. What might be useful, especially if combined with increased emphasis on examiner feedback, would be to include them in the practical test to emphasise their importance. There would be implications here for examiner training, and the current reluctance of instructors to participate in feedback sessions would need to be addressed – as would the receptiveness of candidates themselves to such feedback. In fact, directly influencing driver training to emphasise the risks of such behaviour and to change attitudes towards it may be at least as valuable as modifying the driving test itself; combining the two approaches seems likely to be the most effective option. Measures to penalise violational behaviour during early post-test driving could also be useful. An example of such a measure is the provision under the Road Safety (New Drivers) Act 1995 whereby drivers who accumulate six penalty points within two years of passing the test have their licences revoked.

The EC project TEST (Towards European Standards for Testing) examined the possibilities for bringing these aspects into practical driving tests (Baughan et al., 2005). That project, which involved audits of a large sample of driving tests and surveys of examiners in six countries also provides quantitative data on the test duration and content that examiners and auditors believe is needed for a robust assessment of candidates’ driving.

A further aspect of driving not covered by the current test content is solo-driving decision-making such as choosing where and how to conduct manoeuvres, dealing with missed turnings, route choice, and driving towards a specified destination by using direction signs. Given that novice drivers have to make such decisions immediately after passing the test, there is an argument for representing
them in the test. The benefits of so doing, and the question of whether they would outweigh the difficulties and costs, require further examination. However, some other countries’ driving tests do cover such aspects.

**Administration of test technical standards**

The Driving Standards Agency gives much attention to devising and selecting test routes, and to training examiners and monitoring their performance. Only people aged 25 and with extensive recent driving experience may become examiners. They have to pass a stringent driving test, lasting over an hour, followed by a panel interview. They then attend a residential four-week pass-fail course at DSA’s training centre. This features continuous assessments, progress checks, and final tests of theoretical knowledge, driving ability, and ability to conduct and assess a driving test to a consistent high standard1. The probationary examiner is then attached to a driving test centre where his or her performance is further monitored and developed. Once in post, examiners are further monitored, and attend regular refresher training. Examiners’ written test reports are monitored regularly by the centre manager and frequently by DSA senior managers. Line managers regularly accompany each examiner on test, assessing and marking the candidate and then comparing notes with the examiner afterwards. Results of these observed tests show very good agreement between examiners on test outcome when they assess the same test. Senior examiners from DSA headquarters who are independent from the examiner’s line management chain regularly carry out audit inspections of test centres, in which they accompany examiners on test and check documentation and procedures. Management information gives a regular analysis of the tests conducted by each examiner and enables comparisons between examiners and between test centres. At least annually, managers carry out a detailed analysis of the tests conducted by each examiner, examining usage of test routes, choice of manoeuvres, number of faults of each type, and number of times the examiner has to take action to avoid a dangerous situation.

This is a very well-developed system for establishing and maintaining test standards. Nevertheless, there are variations in pass rate between routes, examiners, centres and regions, – as shown in Section 5 of the report. These do not necessarily imply that there are undesirable variations in test standards, but a better understanding of the factors that affect pass-rates is desirable.

Clearly, differences in test routes within and between centres have a bearing on test standardisation. The report’s findings on the link between pass rate and free driving time, roundabouts and signal-controlled pedestrian crossings are relevant here, though since the survey was undertaken DSA has made some changes to test routes to improve coverage of road and traffic conditions.

Full route standardisation would be impossible to achieve because suitable routes are not available. Some

---

1 A theory test, including a hazard perception test, now forms part of the initial selection process for examiner training.

Drivers taking the test while it was raining had a pass rate about three percentage points lower than those taking the test in the dry. A further examination of the reasons for

Drivers taking the test while it was raining had a pass rate about three percentage points lower than those taking the test in the dry. A further examination of the reasons for
this might suggest ways of improving candidates’ ability to deal with such conditions.

**The test marking system**
The current marking system was generally liked by examiners and instructors who participated in the project. It achieves good agreement between examiners who observe the same test, and is ‘objective’ in the sense that it allows examiners to explain a failure decision by pointing to a specific event. It also represents a considerable investment by DSA in the training of examiners, and so is not something to be changed lightly. The system is, however, inherently limited in its ability to estimate a candidate’s underlying rates of making serious or dangerous faults. This tends to lead to low test–retest reliability, and may reduce candidates’ perceived control over test outcome, hence discouraging them from taking more training and practice before coming for test. In principle, it should be possible to improve the performance of the test as a measuring instrument by increasing the number of errors observed — ie increasing test duration or basing the test on less severe (more frequent) faults. However, the benefits of doing this are uncertain, and there would be practical difficulties. For example, as the report shows, a much longer test would be needed to make any substantial difference to its ability to estimate a candidate’s underlying frequency of making serious or dangerous errors. This would lead to a substantial reduction in pass rate, at least initially, unless the current failure criterion (ie a single serious or dangerous fault) were to be relaxed.

The fact that the test focuses on failures of performance rather than on assessing good performance, is also sometimes held to be a disadvantage.

Although the current marking system does have some limitations, what is not clear at present is (a) how important they are, and (b) whether alternative types of test are any better — ie whether they successfully solve the limitations without introducing new ones. Alternative types of marking system — eg based on examiner judgements of competence (as used in Sweden and Western Australia), or based on the proportion of times a candidate performs an action correctly (as used in some Australian and North American states) would need to be carefully evaluated before any decision on implementation was made.

Of course, any limitations of the current marking system would be reduced in importance if other measures were brought in to supplement the test as a way of influencing driver training and practice. Such measures are discussed later in this executive summary and in Section 10 of the report.

**Predictive validity**
It is impossible to measure the ability of a fully implemented driver licensing test to predict candidates’ future accidents, since people are not permitted to drive unsupervised until they have passed the test. If its ability to predict future accidents could be measured, there are many reasons to expect it to be rather low for real candidates, even for a test that is effective in maintaining safety standards. An indication of predictive validity can be obtained by examining whether test passers who perform well during the test have fewer accidents than passers who perform relatively badly. The project investigated the relation between driving faults (ie the less serious faults) on test and accidents during the first six months of post-test driving. The results indicate that once the effects of age, mileage driven, and driving in the dark have been adjusted for, there is a predictive relationship between accidents and number of driving faults. It therefore appears that high-faults candidates do tend to be intrinsically less safe as drivers than candidates who make few faults on test. In principle, the ability of the test to identify relatively unsafe drivers could be improved by tightening the failure criterion for driving faults. However, the argument for this is finely balanced since high faults drivers tend to do less driving than low faults drivers during the first 6 months after passing the test, and this weakens or removes the relation between faults and actual accidents. Note that the survey was only able to study the accident liabilities of people making less than 16 driving faults, since those making 16 or more failed the test.

**Reliability**
The between-examiner reliability of the test appears to be good, as evidenced by the results of routine check tests conducted by DSA. This form of reliability — ie the degree of agreement between pairs of examiners observing the same test — is the one most under DSA’s control via selection, training and monitoring of examiners. Differences between tests associated with routes, time of day, traffic levels, chance events and weather conditions will introduce another source of unreliability. The test marking system, in which a single serious or dangerous fault leads to failure, makes it vulnerable to such unavoidable between-test differences. Such effects will tend to reduce test–retest reliability — ie the level of agreement between pairs of tests conducted on the same candidates.

A major influence on measured test–retest reliability is the fact that candidates choose when they come for test. Many appear to do so at a time when their own level of competence is such that they have only a moderate probability of passing. Their performance is not consistent: their underlying rate of making serious or dangerous faults means that they may or may not make one in any given test. Again, this characteristic can be addressed to a certain extent in a driving test: the test can, in principle at least, be adjusted so that it is able to estimate more precisely a candidate’s underlying rate of making faults. However, even if the precision of the test were to be improved, candidates might still choose to come for test when they are in that range of competence within which the test is not able to discriminate well.

**The test standard**
The fact that candidates tend to come for test at a level of competence that gives them only a moderate probability of passing is of fundamental importance in understanding how the practical driving test operates. It means that, in
terms of their underlying rate of making serious or dangerous faults, many people who pass the test are no better than those who fail. This is not necessarily a problem for road safety, and should not in itself be regarded as a deficiency of the driving test. What it does mean is that the test needs to be sufficiently demanding that candidates who have only a moderate probability of passing are nevertheless good enough to be permitted to move to the next stage of licensing. In Britain this involves being permitted to drive unsupervised with one probationary restriction.

If we judge that people with only a moderate probability of passing the test are not competent enough to drive unsupervised, then several types of solution are possible:

a Increase the severity of the test. One way of achieving this would be to increase its duration while keeping the failure criterion unchanged. A problem here is uncertainty as to how effective the new, more severe, test would be in inducing learners to reach higher standards. This is likely to depend on the degree to which learners are able to reduce their error rates by taking reasonable amounts of training and practice.

b Ensure that drivers who pass the test do not immediately go on to drive unsupervised and unrestricted.

c Persuade learner drivers to delay coming for test until their probability of passing, and therefore their competence, is higher. Pass-rates would increase, the standard of novice drivers would improve, the number of re-tests would decrease and test–retest reliability would increase. The degree to which this could be achieved in practice is not at present known.

In summary, it is possible to argue that the severity of the current test is too low, as evidenced by the current novice driver safety problem. However, driver competence and safety might best be improved by means of mechanisms outside the test, rather than by raising the test standard.

Pass-rate and candidates’ competence
The pass rate in 2002/3 was 43-44%, and it appears that a substantial proportion of candidates have individual probabilities of passing of around this value. It might be argued that this is neither good nor bad, and that if candidates wish to risk repeated test failures, they have the right to do so. However, in view of the benefits of increasing the probability of passing the test (see (c) above), it is useful to consider how the current pass rate may have arisen and what might be done to increase it.

An obvious possible explanation is that the rather low pass rate results from candidates attempting to minimise the total cost of obtaining a driving licence by coming for test too early in the hope that they will pass. No doubt this is one component. The report also indicates that instructors’ optimism about the chances of a candidate passing may be another. Expectations of parents and competition with friends may also play a part. However, the evidence presented in the report indicates that there are other factors at work that might be addressed by changes to the test itself. These include candidates’ beliefs that their moderate probability of passing arises not so much from limitations of their own driving, but from factors outside their control. The generality of this finding, the reasons for it, and possible solutions, need further investigation. If it were possible to increase candidates’ perceived control over test outcome, they might then choose to delay coming for test until they had a higher probability of passing, especially if the consequences of failure were increased by (for example) increasing the cost of retests or the compulsory minimum delay between test and retest.

However, research would be needed to investigate the extent to which candidates are able to reduce their rates of making serious and dangerous faults by taking increased amounts of training and practice. This would extend previous DfT research on the relationship between error rates, training and practice.

Again, other interventions designed to increase the amount of training and practice accumulated by learner drivers may be more effective at increasing the pass rate than changes associated with the test itself.

Feedback from examiner to candidate or instructor
The driving test is an opportunity for candidates, instructors and supervisory drivers to receive feedback from the examiner. This could cover the candidates’ competence and limitations, and the importance of motivational/attitudinal influences on driving. Only a short feedback session at the end of the test is included at present. Also, despite encouragement from DSA, few instructors choose to be present in the car during the test, or to listen to the examiner’s feedback to the candidate after the test. The reasons for the apparent reluctance merit further examination. Instructors appear, from the evidence presented in the report, to be optimistic about candidates’ probabilities of passing. This, and the fact that it is desirable for candidates (especially those who pass the test) to appreciate their own limitations, means there is a good argument for increasing the emphasis on feedback.

Candidates may not be receptive to detailed feedback immediately after the test, and the provision of instructors’ time, examiners’ time and debriefing facilities would be expensive to organise. Nevertheless, a lot of emphasis is given to test feedback in some other countries, and it would seem worth exploring further in Britain.

The licensing system
Changing the driving test should not be seen as the only, or necessarily the best, way of improving novice driver safety. The licensing system itself can be modified to encourage or require learner drivers to take more and better training and practice before driving solo, and to influence behaviour and exposure to risk during the early months of solo driving. These possibilities are discussed in chapter 10 of the present report and in more detail by Baughan and Simpson (2002). Such changes have been widely and successfully applied in other countries, and
have been considered in Britain in the Department of Transport’s consultation document ‘Introducing a more structured approach to learning to drive’ (DTLR, 2002). They include:

**Increasing the amount of driving experience accumulated by learner drivers**

Increasing the amount of driving experience accumulated before solo driving has been shown to reduce novice drivers’ accident risk. If such increases in experience could be achieved in Britain, there would almost certainly be an improvement in novice driver safety. Possible mechanisms include increasing the minimum age for holding a full licence, reducing the minimum age for starting to learn to drive on the road, and/or introducing a minimum learning period. Increases in pre-solo experience can also be achieved by specifying suitably high minimum amounts, and requiring supervisors to certify that the requirement has been met. Advisory minimum targets for pre-solo experience also seem likely to be helpful. A logbook for learner drivers, such as the voluntary logbook introduced by DSA, should be useful in this regard.

**Night-time restrictions**

There is evidence that these can be very effective at reducing night-time accidents during the months covered by the restricted licence – though clearly this will depend on the level of enforcement and penalties. If night-time restrictions were to be considered in Britain, the likely benefits would need to be compared with the effects on employment and mobility, which could be minimised by choice of curfew period and by exemptions for particular types of journey.

**Passenger restrictions**

Research has shown that the presence of young passengers can adversely influence the behaviour of young drivers. The passengers also suffer from the novice drivers’ already high accident liability. A passenger restriction for drivers when they first begin driving unsupervised could therefore have safety benefits. In Britain, there is currently no requirement for drivers/passengers to carry evidence of age or, indeed, a driving licence. This might need to be changed to facilitate enforcement, though parental influence will be important for many drivers. The possibility of young people making more car journeys as car drivers rather than passengers, thus increasing the total number of journeys, would need to be taken into account before a decision were made on whether to introduce passenger restrictions. So, too, would social effects such as reduced mobility, and impact on schemes to reduce drink-driving by promoting the idea of ‘nominated drivers’.

**Increasing penalties for traffic violations**

The association between traffic violations and accident liability, as well as other considerations, suggest that reducing novice drivers’ propensity to commit traffic violations would be beneficial to safety. Many licensing systems contain elements that require a period of conviction-free driving before moving to the next stage, or enable licence sanctions to be introduced at a lower threshold than is the case for fully licensed drivers. The British system of licence revocation for drivers who accumulate six penalty points within two years of passing their driving test is an example of such a provision. Such measures can be seen as a way of maintaining a supervisory influence on novice drivers during their period of early solo driving. There is, as yet, little evidence on their effectiveness, but they are relatively simple to introduce and are attractive in that they seek to address the motivational components of novice driver safety.

**Improving training and education**

There are many possible reasons for the general lack of evidence that increased driver training improves safety, and these have been discussed extensively in the literature. A number of themes have emerged, which offer the hope of improving the effectiveness of training. For example, current training of learner drivers tends to concentrate on car-control skills This may increase drivers’ confidence in their ability to cope in difficult situations, and neglect higher order skills associated with hazard perception and decision making – which was one of the reasons for introducing the hazard perception test into the theory test in November 2002. Factors associated with drivers’ attitudes, goals and motivations also appear to be important influences on behaviour, but are not well covered in conventional training. Training that addresses these areas – at least in terms of raising people’s awareness of their influence of attitudes and goals on driving – may well prove to be effective in improving safety. Hatakka et al. (2002) provide a discussion of these issues. Developing and evaluating such training is now an important task.

**Reduced alcohol limits for novice drivers**

Given the problems of alcohol related accidents amongst novice drivers, imposing lower limits on young or novice drivers is likely to bring benefits and may also instill safer drink/driving habits even after the restricted period ends. However, in Britain, enforcement of a differential limit for novice drivers would be difficult in the absence of a requirement to carry licences or identity cards. Also, drink-driving is more prevalent amongst those in their early 20s than it is amongst teenage drivers. The view has sometimes been expressed that it may therefore be counterproductive to have a lower limit for novices, who would then see the limit raised just as they moved into the group in which the drink-driving problem peaks.

**Driver and passenger identification**

Enforcement of several of the measures listed above would be difficult unless novice drivers, and possibly their passengers, are required to carry identification.

**Probationary licences and exit tests**

The British licensing system already includes a two-year probationary period after the practical driving test has been passed. Drivers who reach the end of the period without
accumulating six or more penalty points automatically exit from the probationary phase. It would be possible to make this conditional on passing a further test, as is done in several graduated licensing systems. However, this report concludes that adding an exit test to the end of the current two-year probationary period does not have much to recommend it in Britain at present, although this conclusion might change if more severe risk-reducing restrictions were to be imposed during the probationary phase.

It could well be useful to alter the name of the post-L-test licence to (say) the probationary licence. This should make novice drivers more aware of probationary conditions and facilitate enforcement of the any restrictions imposed during the probationary period. However, there would be costs associated with issuing a further licence at the end of the probationary period. In fact, it appears that EC legislation would not permit the introduction of a post-test probationary licence. But it would be possible to issue the first full licence with a special code for a probationary period, and issue a full ‘till age 70’ licence after this period.

Conclusions and recommendations

**Tightening the failure criterion for ‘driving faults’**

There is a valid argument for tightening the failure criterion for ‘driving faults’ from its current limit of 16. Such a change would cause a group whose driving is intrinsically less safe than that of others to fail the test, and therefore to have more training and practice before driving unsupervised. However, this argument is finely balanced: in practice this group of people tend to do less driving than ‘lower-faults’ drivers, at least in the first six months of post-test driving. So, although their driving is intrinsically less safe, their actual number of accidents is similar to that of lower faults drivers.

**Persuading people to delay coming for test**

It is desirable to find ways of persuading people to accumulate more driving experience before coming for test. This would reduce post-test accident liability, and should also increase the test pass-rate. However, increasing test fees and delaying re-tests might not be effective in persuading L-drivers to delay coming for test in order to improve their probability of passing. More needs to be known about the ability of candidates to reduce their error-rates (and increase their test pass-rates) by taking additional training or practice. Learner drivers’ perceptions of their ability to influence test outcome in this way also need to be better understood. The apparent optimism of both candidates and instructors regarding the probability of passing the driving test also merits further examination.

**The content of the practical driving test**

The possibility of broadening the test to include independent driving skills such as choosing where and how to conduct manoeuvres, dealing with missed turnings, finding a route using direction signs, and route planning, merits consideration. The potential for strengthening the assessment of hazard perception during the practical driving test, in order to complement the computerised hazard perception test that is now included in the theory test, is worth exploring further. The growing realisation of the importance of drivers’ attitudes and goals as influences on behaviour (Hatakka et al., 2002), and the need for drivers to be aware of these influences and how to recognise and control them has implications for both driver training and testing. It may not be feasible to incorporate such aspects into the pass-fail criteria for the test but they could in principle be incorporated in the practical test to emphasise their importance (Baughan et al., 2005).

**The test marking system**

Tests of the types used in Sweden, the Netherlands and Western Australia place less reliance on the making and recording of individual errors than the British test, and more reliance on examiners’ judgements of competence. If further improvements to driver testing are sought, it would be useful to investigate these tests alongside the British test to assess whether they offer benefits, and whether these are likely to be sufficiently great to justify the investment that would be needed if elements of such tests were to be introduced in Britain.

**Consistency of test standards**

Between-centre differences in pass-rates merit further examination to discover whether they represent undesirable differences in route/traffic demands or assessment standards, or whether they are explained by differences in the competence of candidates coming forward for test, perhaps associated with differences in access to training and supervised practice. A study of the extent to which drivers restrict their early driving to the area where they take their test would be instructive in helping to assess the importance of regional differences in test conditions.

If there are identifiable types of test centre at which candidates come for test particularly poorly prepared, then low pass-rates at such centres would imply that the test is doing a useful job in screening out such candidates. However, policies to improve candidates’ preparation at these centres may be needed.

The effect of the choice of reversing manoeuvres on pass rate is complex, involving interactions between difficulty of the manoeuvre, time of day of test, and possibly other variables. Nevertheless, the rather large differences between the manoeuvres in terms of the probability of making a serious or dangerous fault mean that this area merits further examination.
Other aspects of the training/testing/licensing system

Changing the driving test should not be seen as the only, or necessarily the best, way of improving novice driver safety. The licensing system itself can be modified to encourage or require learner drivers to take more and better training and practice before driving solo, and to influence their behaviour and exposure to risk during the early months of solo driving. Such changes have been widely and successfully applied in other countries and have been considered in Britain in the Department for Transport’s consultation document ‘Introducing a More Structured Approach to Learning to Drive’ (DTLR, 2002).

References


1 Introduction

The accident liability of novice drivers decreases very sharply during the first six months of post-test driving, and continues to decline markedly over the next 2 - 3 years. Much of the improvement is known to be associated with experience rather than age (eg Maycock, 1991; Maycock, 2002). This implies that safety could be improved by (a) enhancing the learning process, (b) preventing people from driving unsupervised until it has taken effect, or (c) influencing behaviour or reducing exposure to risk during the early months of unsupervised driving to counter the effects of inexperience and immaturity.

One approach here is to modify the driving test to improve the training and experience accumulated by learner drivers, and screen-out drivers who have not yet reached a standard acceptable for solo driving. The Department for Transport (DfT) commissioned TRL to assess the scope for such improvements to the practical driving test for car drivers.

Studies were undertaken of the consistency of candidates’ performance at the time they come for test, the way in which candidates decide when to come for test, the literature on novice driver accidents and its implications for driver testing, the influence of route and other characteristics on the pass rate, and the relation between numbers and types of less serious faults made during the test by people who pass it, and their subsequent accidents.

In Britain, the driving test is the main tool for inducing learner drivers to build up training and experience before driving solo. This is asking a lot of a driving test. In addition, it is difficult for a test to include in its pass/fail criterion those variables that govern the discrepancy between supervised driving performance during the test, and the subsequent unsupervised driving behaviour of novice drivers. It is therefore necessary to consider whether other changes to the training/testing/licensing system would be beneficial. Graduated licensing systems used in other countries aim to provide a staged progression from initial learning to unrestricted solo driving by means of measures designed to restrict exposure during early driving, exert a supervisory influence over driver behaviour during the first part of a driver’s solo driving career or improve the level of training and experience accumulated before driving solo without restrictions. The project included a review of graduated licensing and related systems in other countries with a view to identifying measures that might be effective in Britain.

This report summarises the main elements of the project. Detailed information on some of the project subtasks is published elsewhere – (eg Maycock, 2002; Baughan and Simpson, 2002).

2 Assessing driving tests

2.1 The British driver licensing system

A driving test needs to be assessed in the context of the licensing system in which it operates. In Britain, the licensing system for car drivers allows learners to obtain a ‘provisional’ licence from age 17. This would be termed a ‘learner permit’ in some other countries: it allows the learner to drive (not on a motorway) whilst supervised by a driving instructor or another driver who has held for at least three years a valid, full licence for the category of vehicle being driven, and is at least 21 years old. Learner drivers are not required to take any specified amount of training or practice – an important point when it comes to understanding the function of the driving test in Britain. A theory test and a practical test can be taken from age 17, the only restriction being that the theory test has to be passed before the practical test is booked. There is no minimum learning period and if both tests are passed a full driving licence is issued. There are no special restrictions on where or how a newly qualified driver may drive, though there is an element of probation: drivers who accumulate six or more penalty points within two years of passing the test have their licences revoked.

2.2 Objectives and functions of the driving test

In a project to review the driving test and suggest improvements, it is important to understand the objectives of the test and how it seeks to achieve them.

The objectives clearly include achieving acceptable levels of safety. Competence to make full use of a car in a socially responsible way would also seem to be important, as would the fairness and efficiency of the test.

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/or practice and take the test again, repeating the process until they pass. Thus it can be seen that 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 unsupervised. This is especially so in a training, testing and licensing system like Britain’s, that does not require learner drivers to take any particular training or supervised practice. This crucial function of driving tests is well recognised (eg Waller et al., 1978; Macdonald, 1988; McKnight, 1992; Mynttinen, 1996; Baughan, 1998). It has important implications both for the design of driving tests, and for any study that aims to assess and improve them.

Figure 1 summarises the main objectives and functions of the British practical driving test.

2.3 Validity, reliability and pass-rates

2.3.1 Reliability and validity

Fundamental in the assessment of any test are the concepts of reliability and validity.
The validity of a test is usually defined as the degree to which the test measures what it purports to measure. For example, given that the main objective of the driving test is to do with road safety, a test would have **predictive validity** if it were able to distinguish between candidates who would go on to have high and low accident liability. Other safety-related variables that a valid test might be expected to predict could, in principle, include near-accidents, performance in special driving assessments, or the meeting of a defined list of driving competences. The ability of a test to discriminate between different groups of drivers known to have different levels of competence or safety – eg novice and experienced drivers – could also be investigated. Of course, a criterion variable other than accidents would need to have its own validity established as an indicator of safety or competence. Section 9 describes a study to investigate an aspect of predictive validity – ie whether the number ‘driving faults’ (the less serious faults assessed in the driving test) predicts the accident liability of people who pass the test.

Other types of validity include **content validity** and what may be termed **consequential validity**. Content validity concerns the coverage of the test. A test would be judged to have good content validity if it covered all aspects of driving known or judged to be relevant to its objectives. Content validity is closely linked to the concept of driver competence since establishing a set of desired competencies would lead directly to an outline description of test and training coverage. A further aspect of content validity is the quality of the testing – for example, whether it complies with established knowledge and good practice in giving feedback, having a marking system that minimises the influence of chance events and promoting consistency of standards. Consequential validity could be defined as the extent to which a test achieves what it sets out to achieve. Thus a driving test would have good consequential validity if, when introduced into a testing/training/licensing system, it influenced the amount and quality of training and practice undertaken by learner drivers so as to achieve acceptable levels of safety and competence. The consequential validity of a new or modified driving test could, in principle at least, be assessed by monitoring the effect on accidents or driver competence of introducing the test into a licensing system.

The reliability of a test is its ability to produce consistent results: a test is considered to be perfectly reliable if it would produce the same results when repeated under identical circumstances. One way of assessing reliability would be to compare the markings given by two examiners observing the same tests. Another is to give a sample of candidates a test, followed by a re-test, and compare the outcomes. A study of the test–retest reliability of the driving test is reported in Section 6.

There are special characteristics of driving tests that make it difficult to assess predictive validity. For example:

- **a** only a fully implemented test will influence the training and practice accumulated by learner drivers.
- **b** people who fail a fully implemented test are excluded from unsupervised driving and therefore from any follow-up study of accidents during unsupervised driving.
- **c** there is a difference between how drivers are able to perform (eg while being tested) and how they then choose to behave while driving unaccompanied. This will reduce further the ability of driving test results to predict accident liability. Section 3.3.9 discusses the scope for dealing with this problem in the driving test itself, and Section 10 examines how probationary licensing restrictions may be used in an attempt to influence novice drivers post-test behaviour.

Baughan (1998) and Appendix A discuss some of these issues in more detail, and show why a test may have low

---

**Figure 1** Objectives and functions of driving tests

<table>
<thead>
<tr>
<th>OVERALL OBJECTIVES:</th>
</tr>
</thead>
<tbody>
<tr>
<td>To help the training/testing/licensing system to achieve:</td>
</tr>
<tr>
<td>- Acceptable levels of safety</td>
</tr>
<tr>
<td>- Acceptable competence (to make full use of car in a socially responsible way)</td>
</tr>
<tr>
<td>- Efficiency</td>
</tr>
<tr>
<td>- Fairness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FUNCTIONS (HOW A TEST SEeks TO MEET ITS OBJECTIVES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Selection:</td>
</tr>
<tr>
<td>Screen out drivers who are still at a stage where they pose an unacceptable threat to themselves or others</td>
</tr>
<tr>
<td>In influencing training and practice: (this is the main function)</td>
</tr>
<tr>
<td>(a) Further training and practice for screened drivers</td>
</tr>
<tr>
<td>(b) Test syllabus and standards influence content, standards and amount of training and practice</td>
</tr>
<tr>
<td>(c) Training content of the test itself</td>
</tr>
</tbody>
</table>

*Note: (b) and (c) reduce need for (a)*
reliability and predictive validity for candidates who actually come forward for test, but may still induce good training and practice and thus have high consequential validity.

2.3.2 Pass-rate

Pass-rates for practical driving tests in different countries vary from 99% to less than 20%, with many countries having a rate of around 50%. In Britain, the rate in 2002/3 was 43-44%. A low or moderate pass rate is sometimes taken to indicate that a test is unjustifiably severe, or that a high proportion of candidates are not ready for the test in the sense that they have not prepared sufficiently for it and have not reached the desired level of driving competence. Likewise, a very high pass rate for a test is sometimes taken to indicate it is ineffective at identifying poor candidates – ie that it is too easy, because it ‘passes everybody’. However, the test pass-rate, and the concept of being ‘ready for test’ need to be considered alongside the related issues of test reliability, test severity, and individual candidates’ probabilities of passing (Baughan, 1998). For example, to take two extreme alternatives, a test pass rate of 40% could indicate:

- 40% of candidates each have a very high probability of passing and 60% of candidates are not ready for test, each having a very low probability of passing.

or:

- All candidates come for test having reached a similar level of competence, such that they each have a probability of passing of 0.4. (This would imply that for these candidates, the test has low test–retest reliability.) All these candidates will be in an equal state of readiness to take the test and, if the test standard is set high enough, will have reached an adequate level of competence to be allowed to drive unsupervised. However, the same level of competence could be achieved by a somewhat less severe test if candidates could be persuaded not to apply for it until their probability of passing (and thus the overall pass rate) is higher. Section 6 explores such issues further.

There are a number of reasons why it would be better to achieve a given level of competence through candidates having a high probability of passing a less severe test rather than a low probability of passing a more severe one. One is that a low pass-rate may tend to put strains on the relations between candidate, examiner and instructor. Learners who fail may feel tempted to blame either instructor or examiner, especially if feedback is not adequate or if instructor does not agree with this feedback. Instructors themselves may tend to be dissatisfied with the performance and the judgement of examiners if they repeatedly see candidates fail, whom they judge to be ready for test. Examiners may come to believe that instructors are not performing well and succumbing too early to learners’ wishes to apply for the test. Furthermore, examiners have to cope with the sometimes strong emotional reactions from learners that may follow upon a failure notice – reactions that make it difficult to deliver effective feedback.

In addition, a high failure rate means that there are large numbers of driving tests being conducted which would be unnecessary if candidates could be dissuaded from coming for test until they had a high probability of passing.

A regime in which a sizeable proportion of candidates are coming forward for test with only moderate probabilities of passing would mean that, for these candidates, it would be largely a matter of chance whether they pass or fail any given test. This seems intuitively undesirable in that it may foster the idea that the test is a lottery and not something to be prepared for seriously.

2.3.3 Causes of low pass-rates

The question arises what causes a low or moderate pass rate. There are several possibilities – for example:

- Learners may realise that in their current state of competence they have only a moderate probability of passing, but may choose to come for test anyway – perhaps because they think it is worth having an early try at taking the test, or because they think that it would be difficult to increase their probability of passing.

- Learners may over-estimate their probability of passing. In principle, this could be because instructors’ own estimates are optimistic, or because they are reluctant to tell candidates the true situation – perhaps fearing that to do so would look as though they were trying to persuade the candidate to purchase more lessons. Again, instructors may themselves think that there is not much more that can be done to give candidates a higher probability of passing.

Sections 4, 6 and 7 present further information on these possibilities.

2.4 Conclusions on assessing driving tests

1 A driving test needs to be assessed in the context of the licensing system in which it operates. In Britain, the licensing system currently relies on the test as the main mechanism for influencing the training and practice accumulated by learner drivers.

2 The main criterion that the driving test needs to satisfy is that it should be successful in influencing training and practice, such that drivers reach an acceptable level of competence and safety before being allowed to drive unsupervised. That is, the test should have good ‘consequential validity’. However, this consequential validity will tend to act to reduce the range of candidates’ competence and accident liability, and thereby reduce or remove test–retest reliability and predictive validity for candidates who come forward for test. This tendency will be exacerbated if candidates are motivated to come forward for test when they have only a moderate probability of passing.

3 There are methodological difficulties with assessing the predictive validity of driving tests, including the facts that (a) only a fully implemented test will influence the
training and practice accumulated by learner drivers, and (b) people who fail a fully implemented test are excluded from solo driving and therefore from any accident follow-up study.

4 Pass-rates for tests do not necessarily imply anything about the competence, or probability of passing, of people who do pass the test. In principle, those who do pass might have a very high probability of passing and be very competent. On the other hand, passers might not be very different from failers, and might have only a moderate probability of passing. But even candidates with only a moderate probability of passing may have acceptable competence if the test standard is set high enough.

5 A given level of competence could be achieved through candidates having a moderate probability of passing a severe test, or a high probability of passing a less severe test. The latter is preferable since, for example, a higher pass rate reduces the incidence of repeated negative experiences for candidates, instructors and examiners – experiences that may not be helpful to the process of driver training. A high pass rate also reduces the total number of tests and retests needed.

6 It may not be possible to achieve high test–retest reliability and high pass-rates by means of changes to the test alone, since candidates may continue to come forward for test when they have only a moderate probability of passing. To increase pass-rates it may be necessary to find other ways of encouraging learners to reach higher levels of competence before coming for test. The reasons for the moderate pass rate for the British practical driving test are worthy of investigation.

7 The influence of motivational and attitudinal factors mean that a driving test cannot necessarily be expected to give a good prediction of unsupervised driving behaviour and future safety.

8 Assessing the predictive validity of an established driving test is not straightforward and, short of implementing a change and trying to monitor accidents as training adapts, the same applies to a possible improvement to the test. Heavy reliance on content validity – covering the content and quality of the test, and the training and practice that it is likely to induce – is needed when assessing whether changes to a driving test are worthy of full implementation.

9 Assessing an established test:

- A test can have low test–retest reliability and predictive validity for current candidates, but still have a beneficial effect on training – i.e., good consequential validity.
- We would not be justified in rejecting a test, or test items, on the grounds of low predictive validity or low test–retest reliability for current candidates.
- Assessment of predictive validity is problematical because only test passers are allowed to drive unsupervised.

10 Assessing a newly implemented test:

- The initial screening effect may not be a good indicator of the longer term training effect. A test may be successful in identifying (and failing) a group of candidates who would have a higher than average accident liability, but we cannot be sure that training them to pass the test will reduce their accident liability to the average level.

11 Assessing a trial test (ie one that does not affect the granting of a licence):

- Only the screening effect can be assessed directly (i.e., by comparing accident liabilities of passers and failers). The test will not affect training.
- Only people who have passed the current, established, test could be included in the screening study, because only these people would be permitted to drive unsupervised.

3 The British practical driving test: why consider changing it?

3.1 The practical driving test

Before May 1999, the British practical driving test lasted approximately 35 minutes. The examiner chose one of several set routes. The candidate had to execute an emergency stop, a hill start, and two out of three manoeuvres involving reversing (i.e., reverse parking, reversing into a limited opening, and a turn in the road). Driving was continuously monitored for 48 categories of predefined errors, each of which was assessed as a minor, serious, or dangerous fault. Serious or dangerous faults were those judged to involve potential or actual danger, and candidates making one or more such faults failed the test. There was no specified limit on the number of less serious faults a candidate was allowed to make without failing, although examiners could decide that a pattern of repeated less serious faults indicated a serious fault, resulting in failure. There was (and is) no motorway driving, and not necessarily any driving on dual carriageways. Candidates at the end of the test were given a copy of the examiner’s marking sheet, and were offered a short verbal debrief. Instructors were encouraged to attend the debrief (and to ride in the back of the car during the test) but rarely did so.

An early task required of the project was to assess the evidence for and against a number of possible improvements that might be made to the driving test, to see whether any of them could be recommended for adoption without further research. This work, summarised in Section 3.3, identified some changes likely to improve the test and after a public consultation, the Driving Standards Agency introduced the following changes in May 1999:

- Test duration was increased by about seven minutes.
- Coverage of test routes was improved – more routes included sections of higher speed dual and single carriageway roads.
Some examiners believe that a candidate on his or her best behaviour
would pass the test (see Section 4.

The changes involved one in three candidates (rather than all
candidates) being asked to demonstrate an emergency stop. Also, the
reverse parking manoeuvre was now tested the test centre car park
where facilities were available.

3.2 Indications of a possible need for change

3.2.1 Modern traffic conditions

One argument for changing the driving test is that traffic
conditions have altered drastically since it was introduced
in 1935. This is valid in a sense, but since the test observes
driving performance in current traffic conditions it has in
many ways adapted automatically to the changes. There
are some aspects of driving to which the test has clearly
not fully adapted: it does not, for example, cover
motorway driving, and only limited driving at higher
speeds is possible. It may also be that the traffic system has
become so much more complex, with many more
opportunities for conflicts between vehicles, that a more
demanding test of competence is needed.

One way of making a test more demanding would be to
increase its duration while leaving the failure criterion
unchanged. To maintain their current probability of
failing, candidates would have to reduce their underlying
rate of making faults. A longer test would also permit a
wider range of driving conditions to be covered, would
require a candidate to be capable of sustained
performance, and might be better at revealing candidates
whose driving behaviour is likely to deteriorate once they
are allowed to drive unsupervised'. This is discussed
further in Section 6 while Section 5 describes some work
to investigate the conditions covered by test routes, and the
effect these had on pass-rates.

3.2.2 Novice driver accident liability

It is well established that drivers have a very substantially
raised accident liability during their first year or two of
unsupervised driving, and that much of this is associated
with lack of experience rather than their age/immaturity
(see Section 8 and Maycock (2002) for a review). This is
both a problem and an opportunity. It shows that the
present training, testing and licensing system is not as
effective as might be hoped in producing new drivers who
are acceptably safe. It also shows that novice drivers are
learning something valuable during their early solo
driving. This holds out the possibility of achieving
substantial reductions in accidents if ways can be found of
encouraging the learning to take place in a protected
situation – for example, while still being supervised. In
principle, it might be possible to engineer such a
development by changing the driving test itself. Other
ways of achieving it are discussed in Section 10.

3.2.3 Novice driver accident types

The types of accidents that novice drivers are involved in
ought to provide some indication of where their training,
and hence testing, needs to be improved. As part of this
project, Maycock (2002) reviewed the relevant literature
and found that while novice and/or young drivers have
many more accidents of almost every type than older
drivers, they have a particularly high liability for accidents
involving one or more of the following characteristics: no
other vehicle involved; evenings and the early hours of the
morning; weekends; rural roads; single carriageway roads;
bends; overtaking; right turns; speed. The relative
contributions of exposure and driving behaviour to this
pattern of accidents cannot always be ascertained from the
data but, even if an excess of a particular kind of accident
is due to exposure rather than to a deficiency in driving
skill or behaviour, this does not preclude its being
addressed by driver training/testing. The review also found
that young drivers are more likely to be to blame for their
accidents, and tend to be more willing to break speed
limits, engage in threatening behaviour, cut corners, and
drive through amber traffic lights.

These findings suggest that it might be particularly
beneficial to improve learner drivers’ experience of faster
single carriageway roads, manoeuvres like overtaking, and
driving in the dark (though other characteristics of night-
time driving may be more important than darkness – see,
for example Clarke et al. (2002)). To some extent, such
improvements could be induced by changes to the driving
test, though testing overtaking and night driving would be
difficult to arrange in practice.

Further information from Maycock’s 2002 review is
summarised in Section 8.

3.2.4 Research into driver behaviour

A number of research findings imply scope for improving
the driving test:

- The first TRL Cohort Study of novice drivers (Maycock
  and Forsyth, 1997) found relationships between minor
  faults made during the test, and subsequent self-reported
  accidents. This suggested that minor faults may have the
  potential to be used as a failure criterion that identifies
  people with a raised accident risk. If these results could be
  replicated, there might be a good case for failing
  candidates who make specified numbers or patterns of
  minor faults. There would be a reduction in pass rate,
  though this could be expected to recover as candidates
  and the training industry adapted to the new requirements.

- Research on novice driver characteristics and abilities,
  much of it carried out under the DfT behavioural studies
  programme, has identified a number of characteristics
  typical of novice drivers. For example, they perform

---

Footnotes:

1 The changes involved one in three candidates (rather than all
candidates) being asked to demonstrate an emergency stop. Also, the
reverse parking manoeuvre was now tested the test centre car park
where facilities were available.

2 Some examiners believe that a candidate on his or her best behaviour
during the test will find this difficult to sustain if the test is prolonged
– see Section 4.
An important distinction arising from research in traffic psychology is that between a driver’s ability to perform the driving task (eg during a driving test), and how he or she chooses to behave while driving unsupervised. The importance of this distinction is underlined by findings (eg Parker et al., 1995; Maycock and Forsyth, 1997) that people who report making more ‘violations’ (intentional departures from good, safe driving), tend to have more accidents than others.

Intentions to violate formal or informal traffic rules are strongly associated with attitudes (eg Parker et al., 1992). The ability of attitude measurements to identify people who choose to drive badly and have a raised accident liability suggests that it might be useful to measure attitudes as part of the driving test. Difficulties here include the extent to which people are able to fake the attitudes necessary to pass a test, and the question of whether propensity to drive badly would in fact be reduced by educational measures designed to change attitudes. The main benefit of including attitude measurements in the test might well turn out to be to show candidates the importance of ‘bad attitudes’ and rule breaking behaviour as risk-increasing factors, rather than to identify and fail people who have such attitudes.

Related to the previous points is the increasing realisation of the importance of drivers’ life and journey goals in determining behaviour (eg Hatakka et al., 2002). This raises the possibility of including in the driving test items that assess a candidate’s knowledge of such influences and ability to control them (Baughan and Keskinen, 2005).

Over-estimation of skills has been used to explain why some types of driver training lead to an increase in accidents (Gregersen, 1996). Feedback from the driving test examiner could be used to give candidates (both passers and failures) a more realistic view of their own abilities. A feedback session might also provide an opportunity to communicate other information to candidates – including evaluation of attitude scores.

3.2.6 Test marking systems

The probability that a candidate will make a serious or dangerous fault, and thereby fail the current driving test, will depend on the number of opportunities for a fault to occur during the test. This will vary to some extent from test to test (for example, because of varying weather and traffic conditions), introducing an inevitable element of inconsistency or unreliability.

Several alternative types of marking system intended to increase reliability or validity of driving tests have been proposed or are in use in other countries. For example, the idea of standardised routes, or of assessing performance only on those parts of a route that have been standardised, has been introduced in parts of the USA and Australia (eg McKnight, 1992), as has the concept of allowing good performance to cancel out bad performance. Concerns about such tests are discussed in Section 11.4.

Studies have shown that instructors (Hall and West, 1994) and examiners (Quimby et al., 1999) can in some circumstances make judgements about drivers that give some prediction of accident liability. This suggests that the power of the driving test to identify and fail ‘high accident’ drivers might be increased by making use of examiners’ overall judgements of candidates, or by identifying the behavioural observations on which such judgements are based and including these in the test. Tests involving examiners’ overall judgements of skills or competences have been introduced in, for example, Sweden (Mattsson, 1999), the Netherlands, and Western Australia (Drummond, 2000). In principle, this approach might avoid some of the potential disadvantages of tests based on observing errors, but it may also introduce its own difficulties associated with the consistency with which examiners are able to make the necessary judgements. It would also become more difficult for an examiner to refer to a specific recorded event in order to defend failing a candidate. There is also the point that the examiner’s ratings are themselves presumably based to an extent on the errors observed during the test. 

3.2.7 Other considerations

It was recognised at the outset of the project that there might be aspects of the way the test was currently operated by DSA that could be improved. The extent to which test results are influenced by region of country, time of day, examiner and type of route would be relevant here, as would information on regional or between-centre differences in test routes.

3.2.8 European Commission requirements

minimum standards for driving tests. These were, in general, required to be implemented by 30th September 2003. The change required to the practical driving test for cars in Britain was to add vehicle safety checks. This was dealt with by including a brief oral examination at the start of the test (Driving Standards Agency, 2002). From a pool of about 24 standard questions intended to set the learning agenda, examiners now select two at random – normally one ‘tell me’ and one ‘show me’ question – eg:

‘Open the bonnet, identify where you would check the engine oil level and tell me how you would check that the engine has sufficient oil’

‘Show me how you would check that the headlights and tail lights are working’

The European Commission has recently funded a research project ‘TEST’ (Towards European Standards for Testing) in six European countries to investigate whether the driving test provides sufficient opportunities for examiners to make a robust assessment of a candidate’s skills and approach to driving (Baughan et al., 2005). Assessments were made both before and after the implementation of Directive 2000/56/EC.

3.3 Summary of types of change considered for the practical driving test

In the early stages of the study the project team was asked to identify potential improvements to the test and assess whether the available evidence justified their introduction without further research. The likely usefulness and feasibility of research to gain more evidence was also taken into account. Following this review, and after public consultation, DSA introduced a number of changes to the test in May 1999. These are summarised above in Section 3.1. At the same time, other potential improvements were considered and further research tasks for the project were identified. These are summarised below.

3.3.1 Make better use of less serious faults as a failure criterion

In view of the potential indicated by the first Cohort study, it was decided to conduct a study of the relation between less serious test faults (renamed ‘driving faults’ in May 1999) and the accidents experienced by novice drivers in their first six months of driving. Section 9 describes this study.

3.3.2 Add to the range of driving conditions covered

It seems desirable that novice drivers should be competent to drive on dual carriageways and on higher speed non-built-up roads once they are permitted to drive unsupervised. Yet it was possible for a candidate to pass a driving test with little experience of such conditions. There are indications that novice drivers have poor visual scanning on dual carriageways (Underwood et al., 1997). Analyses carried out for the present project (see Section 8) show that young drivers have particularly high accident rates on higher speed single carriageway rural roads. In principle, these findings suggest that the content validity of the test could be improve by broadening the conditions covered – though there would be implications for costs, and the siting of test centres – which in turn would affect the distance some people have to travel to test centres. In May 1999 DSA increased the proportion of test routes that include higher speed dual and single carriageway roads, so that most centres and many routes now have access to such roads. No further work on this aspect was required from the project.

3.3.3 Increase the duration of the test

This would allow a larger sample of behaviour to be observed, increase the amount of free driving that can be assessed, allow a wider coverage of road types, require a more sustained performance from the candidate and make the test more demanding. It would be difficult to justify the cost of a large increase in duration of free driving without further evidence. Even a small increase in duration would be expected to result in an initial drop in pass rate if pass/fail criteria are not adjusted. Pass-rates might be expected to recover as L-drivers and their trainers adapted to the new requirements. In May 1999, DSA increased the duration of the test by about seven minutes. Together with a change to the testing of the emergency stop, examiners were allowed to test one of the reversing manoeuvres at the beginning or end of the test in the test-centre car park where the facilities were available. This yielded about 10-12 minutes extra free driving during the test. The likely effects of further increases in test duration are considered in Section 6 and Appendix B.

3.3.4 Improve the use made of the test as a training opportunity

Improving feedback from examiner to candidate and instructor is potentially an effective way of improving the training of failed candidates, improving instructional standards, and providing both failed and passed candidates with advice about their limitations. A feedback session would also provide an opportunity to communicate other information to candidates – including general training/educational material about novice driving, as well as feedback and evaluation of (for example) attitude scores and examiner assessments. Given the difficulties of incorporating examiner ratings and attitude scores into the pass-fail criterion, examiner feedback may be the best way of using this potentially important information. Leaving aside cost, the main practical problem is the fact that candidates are unlikely to be at their most receptive to feedback at the end of their driving test. Strong emphasis is already given to examiner feedback in some other countries, notably Finland (Keskinen et al., 1996; Mynttinen, 1996) and the Netherlands. A small change to test feedback was made in May 1999 with candidates being given, in addition to a copy of the examiners’ marking sheet and an oral explanation of the main faults, a written explanation to help them understand the markings recorded by the examiner. Further changes to feedback were not pursued as part of this project because of concerns about the receptiveness of candidates to feedback, and the costs associated with provision of rooms and examiner time.
3.3.5 Improve the standard that candidates choose to reach before coming for test
As discussed in Section 2, improving the standard that candidates choose to reach before coming for test could have several potentially beneficial effects: it should increase novice drivers’ competence and safety, increase pass rate and test–retest reliability, and reduce the number of retests needed. The project therefore included a small-scale investigation of the ways learner drivers make decisions about when to come for test and assessed how sensitive these decisions are likely to be to changes in the costs of tests, or the minimum delay between a test and retest. This study is reported in Section 7.

3.3.6 Improve coverage of higher order skills such as hazard perception
Hazard perception elements have always been included in the practical driving test and the opportunities for examiners to observe hazard perception skills will have been increased as a result of the increase in test duration. Also, during the course of the project, DfT, DSA and TRL were developing a computerised hazard perception test to include in the theory test that has to be passed before a candidate is allowed to book a practical driving test (Grayson and Sexton, 2002). Development of techniques to improve the assessment of hazard perception and other higher order skills during the practical driving test was considered by the project Steering Group to be outside the scope of this project.

3.3.7 Examiners’ ratings of performance
As indicated in Section 3.2.6, there could be some potential for making use of such assessments in the driving test. Work would be needed to develop assessment methods that meet the examiners’ desire for objectivity. One possibility here is that of identifying observable driving behaviours that underlie instructors’ and examiners’ judgements. The project steering group decided that this was not to be taken forward during this project, except via work to explore whether a maximum number of less-serious faults should be used as a failure criterion (Section 9).

3.3.8 Other fundamental changes to the marking system
DSA and DfT specified that any new marking system must retain an automatic failure following any serious or dangerous fault. A move towards a marking system which allows instances of good performance to offset some errors was not favoured either, and the project was not asked to pursue these possibilities further, though it was recognised that they may need re-examining at a later date – See Sections 11 and 12.

3.3.9 Add attitudinal items that predict future unsupervised driving behaviour and accidents
At first sight, this offers an attractive way of improving the driving test, since it addresses the problem of predicting how novice drivers will behave once they are free to drive unsupervised. However, it seems unlikely that attitude measurements can contribute to a pass-fail decision, since they are so open to manipulation by the candidate. They may, however, be useful as part of test feedback to candidates. The project steering group decided that this should not be pursued further by the project.

3.3.10 Test the manoeuvres off-road
Finding suitable locations to test reversing manoeuvres on-road occupies valuable driving time and imposes limitations on the area covered by test routes, so there have been suggestions that manoeuvres should instead be tested off-road. In May 1999 DSA introduced a change that required one of the reversing manoeuvres to be conducted in the test centre car park, except at the test centres that do not have access to a suitable car park. Concerns about the impact of the cost of providing sites for off-road manoeuvring led to the decision that the project should not include any further work on this aspect of the test.

3.3.11 Increase the pool from which ‘called for’ manoeuvres may be selected
Test candidates are asked to demonstrate two out of a pool of three manoeuvres involving the use of reverse gear. In principle, the pool could be expanded, even if only two manoeuvres actually have to be demonstrated on test. Such a change could lead to increased experience and practice being accumulated by test candidates, which on its own would be expected to benefit both competence and safety. The change might, however, be thought to penalise unfairly the small group of candidates who, according to the first Cohort Study (Maycock and Forsyth, 1997) have great difficulty in mastering the basic car control skills but have an accident liability lower than that of young males who find it easier to pass the test. There is also the possibility that L-drivers and instructors would increase the attention given to learning to manoeuvre at the expense of tuition and practice in other aspects of driving. The project was not asked to consider the question further.

4 Opinions of examiners and instructors on the test and how it might be improved
4.1 Exploratory interviews with DSA driving examiners
At the outset of the project a small number of interviews were conducted with Driving Standards Agency examiners to obtain some insight into how examiners view the test, and how they think it might be improved. Six examiners were interviewed. Two were male Senior Driving Examiners (Centre Managers), two were male Driving Examiners, one was a male Supervising Examiner (Sector Manager) and the sixth a female Contract Examiner. The purpose of the project was explained, and anonymity assured. The interviews were tape recorded. A ‘qualitative’, exploratory
There was strong support for the idea of lengthening the test, with the following reasons being given:

- Candidates can generally cope for half an hour, but beyond that time the performance or concentration of the poorer candidates diminishes. One reason for this was thought to be that lessons typically involve about half an hour of actual driving, so that learners are trained to be able to cope with this length of time.
- After about half an hour of driving, candidates reveal more about their true level of ability and likely future behaviour. This opinion was based in part on experience with the extended driving tests that have to be passed by people with certain types of driving conviction.
- Lengthening the test would allow a greater range of driving conditions to be covered. In particular, it would bring most test centres within reach of higher speed dual and single carriageway roads.
- Lengthening the test would allow more opportunities for challenging situations to arise, and to assess the candidate’s ability to cope with them.

However, some of these examiners pointed out that lengthening the test would increase the failure rate and be especially discouraging to first time candidates who, it was thought, already had a very low pass-rate. One examiner said that a longer test would be particularly unfair to seventeen year olds, of whom he believed none would pass.

Another examiner believed that testing the manoeuvres off-road would allow greater precision in the test criteria. By using standardised road layouts, the allowable margins of error could be formalised. He believed that at present the manoeuvres are not tough enough.

The idea of removing the manoeuvres entirely from the practical test, and requiring instructors to certify that an acceptable standard had been reached, was very firmly rejected by all interviewees.

### Dual carriageway/higher speed driving

There was a suggestion that some dual carriageway, higher speed, driving should be included in all tests. Joining dual carriageways was said to be a problem for L-test candidates and there was a suggestion that novice drivers have problems after driving on higher speed roads for a period, and then coming off them into urban areas. The problem with conducting a whole test stuck in a traffic jam was also cited in support of the need for more out-of-town driving. It was acknowledged that not all test centres would be suitable for this, though increasing the duration of the test would increase the number of centres that could provide access to dual carriageways.

### Night and motorway driving

Night and motorway driving experience for novice drivers were seen to be highly desirable. However, they were thought to be impossible to include in the driving test. The solution (volunteered or accepted by all interviewees) was to require novice drivers to take further training in these aspects after passing the test. It was felt that instructors could be asked to certify that the training had taken place, and that a suitable standard had been met.

### The marking system

The test marking system was generally thought to be a good one, allowing some discretion to examiners in assessing whether faults were sufficiently serious to require a fail. However, one examiner felt that even with this level of discretion, the test sometimes had to pass people when it was clear to the examiner that they ought not to be allowed to drive solo.

Examiners were clear that they are not looking for the candidate to perform like an experienced driver during the test. In particular, examiners felt that even successful test candidates cannot be expected to have reached the stage where they are able to ‘look ahead’ while driving. One examiner commented on the very small amount of driving time that people have accumulated when they pass their test and emphasised that he was looking only for basic driving skills. Examiners felt the test should identify, and pass, people who have reached that stage of learning where they are now ready to start to gain experience as unsupervised drivers.

One examiner suggested that the marking system could be improved by allowing more detailed information to be recorded on the assessment form and fed back to instructor and candidate – though the danger of an over-complex marking sheet was recognised.
A change that could be considered for the driving test is the inclusion of a rating or ratings of the examiner’s overall judgement about the candidate’s standard of driving. The rationale might be that the current faults-based marking system does not allow examiners to make full use of their ability to judge the candidate’s driving. When this possibility was explored with the examiners, it was clear that they were concerned about the departure from an ‘objective’ system, in which a failure could be justified (to the candidate, or in court) in terms of specific faults observed. It was clear, though, that at least some examiners do consider themselves able to make reasonably robust judgements about the candidate’s driving by about half way through the test. Some examiners also felt they had at least some ability to pick out candidates who, despite performing well during the test, would drive too fast or badly once they were unsupervised. The importance of attitudes and aggression was stressed here.

Commentary techniques

The interview explored examiners’ views on the feasibility and usefulness of asking test candidates for a commentary during part of the test – principally as a way of assessing whether candidates are able to read the road and perceive hazards ahead. The potential value of the commentary technique was recognised. However, the examiners felt that most driving test candidates would find it too difficult to do, partly because the technique is difficult to learn and partly because some candidates are simply not yet able to perceive hazards ahead. Teaching hazard perception and commentary skills to learners was also felt to be difficult. One examiner pointed out that people have to be taught basic driving skills first. Even when it came to teaching hazard perception and commentary, it was felt that instructors would still be having to keep a watch on basic driving because their pupils would still be learner drivers, and that this could pose some difficulties.

Stopping the car to allow candidates to be questioned about hazard awareness

The idea of stopping the car from time to time during the test and discussing with candidates incidents that had just happened – ie as a way of assessing hazard perception and ‘situational awareness’ – was floated by the interviewer. In some ways this was felt to be a good idea, but there was said to be not enough time during the test to do it.  

4.1.2 Other issues

Novice driver safety

Once candidates have passed the test, it was thought that their standard of driving diminishes for a time (eg a year or two). This was mainly because they have to drive in conditions they have not met before, and have not developed the capability of reading the road ahead.

None of the examiners seemed to think that overt risk-seeking, high speed or socially irresponsible driving was a particularly important problem, except for a minority of novice drivers. Rather, they saw the post-test accident problem in terms of (a) lack of anticipatory hazard perception and (b) lack of experience and skill in conditions such as: night driving; driving on, joining or leaving high speed roads; driving in the rain; driving in icy conditions. As mentioned earlier, the preferred solution was to compel novice drivers to take post-test training, which would be signed off by instructors.

The test standard

Although the standard of instruction was criticised, and improving it was seen as highly desirable, changing the test standard was not volunteered as a way of achieving this. For one thing, it was felt that making the test more difficult would reduce the already low first time pass-rate, and be too discouraging to candidates.

There was a suggestion that instructors should be more answerable for the performance of their pupils – eg if six successive candidates presented by an instructor fail the test, then that instructor should be called to a meeting with DSA.

Effect of test conditions

Although it was recognised that some candidates were potentially disadvantaged by having a test in (say) bad weather, a certain amount of discretion was said to be used in an attempt to redress the balance. Also, if a candidate is failing to switch on wipers or demister, the examiner may give a few hints to allow the candidate to correct the situation. In the end, though, a candidate will fail the test if he does not operate wipers or demister when they are needed – and clearly this is a reason for failure that can only occur during some tests conditions.

Traffic conditions obviously vary between tests, and one instructor mentioned that candidates who are stuck in traffic for most of the test have too easy a time of it. The desire to increase the duration of the road test was motivated in part by a desire to expose candidates to a greater variety of traffic conditions.

One examiner though that it was desirable that test conditions vary from region to region. Tests conducted in London are different from those conducted in Wales, but then so too are the real-life driving environments.

Two of the examiners thought that differences between test conditions (within centres) would be minimised by an extension of the test to one hour.

Driving instruction

Some of the interviewees were very critical of the standard of driving instruction. Specific examples of poor instruction were cited. One was candidates not knowing the hand signals, or (mentioned by three interviewees) not knowing where the demister or wiper controls were. This was also mentioned to illustrate the fact that learner drivers may only experience a very restricted set of driving

---

6 A computer-based hazard perception test was introduced into the Theory Test for all learner drivers in November 2002.
7 In fact, DSA already contacts instructors with unusually low pass-rates.
conditions. They may never have met rain or screen-misting by the time they come for test. A further example was the tendency of candidates to switch on the indicators immediately they were asked to drive off or turn, and then to wait for a suitable gap with the indicators flashing.

In general, many candidates were said not to have reached the required standard when they come for their first test. One reason was felt to be that candidates put pressure on instructors to send them for test too soon – and that amongst some younger drivers, there is an element of bravado in passing the test after only a few lessons.

The idea of relying on instructors to certify aspects of training (eg that an acceptable standard had been reached for maneuvres) was felt by some examiners to be completely out of the question.

The examiners felt that driving instruction would be improved if instructors could be persuaded to take more account of the actual faults committed during tests. One way of achieving this is by having suitably detailed information on the marking sheet.

Instructors are encouraged to attend the post-test debriefing, and ride in the car to observe the test itself. This, too, might be expected to improve standards of instruction but the examiners had differing views on the willingness of instructors to seek and act on feedback from the test. Two examiners felt that instructors were generally not interested in the feedback and test-observations now offered by DSA. Instructors were thought to feel threatened by being present when the candidate is being told about faults during the test, since such faults often reflect deficiencies in instruction. A third examiner was not so critical – reporting that some instructors did seem interested in sitting in at the debriefing and observing the test, and that anyway, such feedback had not been offered to instructors until fairly recently. It was said that one reason why examiners like instructors to observe the test is that it reduces the tendency of candidates and instructors to tell each other that a failure was caused by the examiner’s shortcomings rather than their own.

As well as helping instructors improve what they do, feedback was thought to fulfil the useful role of helping them persuade candidates to follow the instructor’s advice.

4.3 Conclusions from the exploratory interviews
These interviews, coming as they did during the early stages of the project, provided a useful source of insight into the driving test as it is seen by examiners, though the very small number of interviews needs to be kept in mind. A number of suggestions for changes to the test were identified, including:
• Increasing the duration of the test.
• Off-road manoeuvres.
• Inclusion of dual carriageways.
• Having instructors sign-off post-test training at night and on motorways.
• Improving the test marking sheet to give better feedback to candidates and instructors.
• Improving co-operation between instructors and examiners, so that more effective feedback can be given. In addition, strong views were expressed about several issues with important implications for the system of driver training, testing and licensing:
• The standard of driving instruction.
• The unacceptability of allowing driving instructors to sign-off manoeuvres.
• The extreme difficulty that most candidates have in ‘looking ahead’.

Further useful insights gained from the six exploratory interviews included the following:
• These examiners believed that they could assess candidates quickly, but rejected the idea of an overall rating of driving competence as being too difficult to justify to candidates.
• They also believed that most candidates are sent for test too soon.
• Examiners are not asking candidates to perform like experienced drivers.

4.2 Group discussions with examiners and instructors
Six group discussions were undertaken, four with examiners and two with instructors. There were seven or eight individuals in each group. All participants were recruited by TRL with co-operation from DSA.

Participants were paid an incentive to attend. Driving instructors were a mix of individual instructors and those from larger driving schools. Examiners were drawn from several test centres in the West London/Thames Valley area. The discussions took place before the May 1999 changes to the driving test were announced. The main findings are summarised below.

4.2.1 Possible improvements to the test
Lengthening the driving test
Two issues frequently arose spontaneously in the discussions, and elicited strong feelings: the length of the driving test, and the role of the manoeuvres. The examiners tended on the whole to feel that it would be an important improvement if the current test could be lengthened. The key aspect they wanted to devote more time to was ‘free driving’ and they contrasted the inadequate time available for this with the time taken up by preparing for and doing the manoeuvres.

Most of the 30 examiners who participated agreed that a small proportion of candidates passing the test were bad drivers who could not be failed since they made no serious or dangerous errors during the test. It was felt that a longer period of free driving would result in such candidates making a serious or dangerous fault, and thus failing the test. For example:
• Whilst an individual might be able to concentrate effectively for 20 or 30 minutes driving, extending this beyond the 30 minutes might reveal a propensity to make errors.

These discussions took place before the 1999 changes to the test, which produced an increase of about 10–12 minutes in free driving time.
Drivers who passed the test by ‘being on their best behaviour’ were more likely to slip into their more natural, spontaneous manner of driving if the test was longer.

Another benefit of a longer test was that it would allow the tests to take in a greater variety of conditions, road and junction types.

Opinions varied as to the ideal length for the driving test: whilst some talked about adding 10 minutes, others wanted to have an hour available (although not necessarily feeling that it should always be used). A further complication was the possibility of reducing the time spent on manoeuvres: it was felt that if this were done it would help by freeing up time for actual driving.

Many of the instructors also favoured extending the test to allow more time on free driving.

Manoeuvres
There was widespread (but not universal) agreement amongst the examiners that the time spent finding appropriate locations and actually doing manoeuvres was disproportionate and would be better spent on free driving.

In some groups the possibility of testing the manoeuvres separately was spontaneously suggested, in others it was introduced by the discussion moderator. Whilst this was seen to have benefits in terms of allowing the main test to focus on ‘real driving’, it was felt to be impractical and, more importantly, unsound. Most (but not all) examiners believed that the manoeuvres were testing not only control of the car, but also observation and hazard perception, and therefore had to be conducted in real life, on-road situations.

The idea of reducing the number of manoeuvres actually called-for on test, but having them selected from a pool that candidates would still have to learn, was favoured by some. There was a suggestion that the emergency stop was suitable for inclusion in this pool.

Instructors tended to hold very similar opinions to those of the examiners on this issue. In one group the suggestion was made that instructors be allowed to sign off their pupils when they had reached a satisfactory level of competence in manoeuvring, but this was a minority view amongst the instructors, who generally did not think such a development was either likely or desirable.

Dual carriageway and higher speed driving
A number of examiners mentioned spontaneously that it would be desirable to include dual carriageways and national speed limit roads in all tests, though there was general acceptance that how easy this is depends on the position of each test centre. It was felt that, in many cases, people could pass their test without it including any driving at higher speeds, or any necessity to demonstrate ability to enter and leave a dual carriageway. In several cases, illustrations were given of otherwise competent drivers who were apparently totally incapable in this respect.

Instructors also tended to feel that dual carriageways should be included in the test, but most accepted that it would be impossible to make this mandatory.

Motorway and night time driving
Most examiners saw motorways as posing a similar problem to dual carriageways and other higher speed roads, albeit perhaps in a more acute form. Again it would be ideal to include motorways in the driving test, but it was accepted as being out of the question to attempt this.

Upon further probing, examiners tended to fall into two camps: some felt that motorways do represent a different set of circumstances and merit consideration as such; others maintained that in terms of genuine driving conditions, there was nothing to differentiate fast dual carriageways from motorways.

Nonetheless, there was very widespread support for action to address the motorway issue. The suggestion that instructors should be allowed to take their pupils onto motorways was generally supported, since it was believed that pupils would not generally be permitted by their instructors to drive on motorways until they had reached a reasonably advanced level. This was contrasted with the current situation where many people who pass their test will then drive on a motorway immediately, without gaining experience or having specific tuition. Some examiners also believed that there should be a requirement for pupils to have a certain number of hours tuition specifically on motorways before they can be issued with a full licence. This could be done before or after the driving test (opinions differed on this), and signed off by an instructor.

Instructors tended to believe that motorways did merit attention, and were unhappy that when they offered post-test lessons on motorway driving (even in some cases free of charge), they were not taken up. Most would like to be able to take learner drivers on motorways, and would be happy to conduct a scheme which involved them in ‘signing off’ their pupils after specific tuition.

As regards driving at night or driving in extreme weather conditions, there was little support for special provisions. Most examiners felt that the skills needed for night driving were not radically different from daytime driving skills, with the only area of concern mentioned occasionally being the driver’s vision (which, it was felt, could not be addressed). The lack of enthusiasm for special attention being paid to this area combined with the practical difficulties tended to stifle discussion on this.

Test standard
When asked whether the current standard needed to pass the test was high or low, a variety of opinions were voiced. Some (largely but not exclusively Senior Driving Examiners) claimed that the standard was low, and supported this by reference to the generally low standard of driving seen on the road and the level of accidents and fatalities. Some also criticised the driving test for being simple and easy to pass without developing any

---

9 The changes introduced in 1999 mean that most test centres do now have access to routes incorporating higher speed roads.

10 In conjunction with the insurance industry, DSA promotes the ‘Pass Plus’ scheme of post-test training. It is in six modules that include night and bad weather driving.
understanding of driving or perception of hazards. However, even the more critical respondents did not go on to suggest any truly radical changes.

Others who were more positive agreed that the test gave no assurance that a driver will have high standards, but argued that that was not the objective. They saw the test as an evaluation of basic competence which should then be built upon. They thought the test had become more difficult over time, as conditions such as traffic volume and amount of aggressive driving have worsened. The current low pass rate was also seen as indicating a high test standard. Whilst these examiners admitted that some drivers passed their tests who probably would not pass in a perfect system, this did not appear to influence their view of the general standard.

The discrepancy between the apparently diametrically opposed opinions found among the examiners seemingly derives from the different points of reference which help shape their views. If examiners think of the test as a central control on driving standards, they may consider that the test standard should be improved in order to reduce accidents. If, on the other hand, examiners regard the test as only one of the influences on general driving standards, and as a ‘screen’ to ensure a reasonable level of competence among new drivers, they are more inclined to feel the current standard is acceptable. They tend to say the difficulty candidates experience in passing the test demonstrates that the standard is high.

Instructors tended to feel that the standard was high but there were exceptions to this. Even among those feeling the standard was high there were no complaints that it was unreasonable or excessive.

The marking system
Examiners were unanimous in their view that the current marking system is very good, with no-one expressing any overall criticism of the system. Whilst more detailed discussion did identify some issues worth examining in more detail, it remains true that from the examiners point of view, the marking system is essentially sound. They were confident in their ability to mark the test objectively, and the marking system appears to contribute to this confidence, which is further boosted by the training and supervision they receive. None of the specific fault types featured in the test were generally thought to be redundant or problematical in any way.

The strengths of the marking system appeared to be that it was practicable to administer in the car, it was simple and clear, and allowed examiners to explain to candidates why they had failed. It also covered the various aspects of driving performance in an efficient way.

The basic approach that a candidate passes unless a specific serious or dangerous fault is made looks clear-cut, but further discussion did show some ‘grey areas’. Examiners, and especially the more senior examiners, did admit to holding off from the decision as to the seriousness of an error whilst continuing to evaluate the candidate. In other cases it was stated that something which appears to be a habitual fault will initially be considered minor, but will be classified as serious after occurring several times. It was stressed, however, that in the majority of cases the result was very clear cut.

The Senior Driving Examiners were quite happy to admit to using (and needing) a little flexibility or judgement in evaluating a test. The less experienced examiners claimed that the whole exercise can be conducted objectively, and justified this by reference to the training which they receive.

One ‘problem’ which was described several times concerned drivers whose overall standard was not felt by the examiners to be acceptable, but who passed the test because they did not make serious or dangerous errors. The opposite situation, in which a basically good driver fails because of one serious or dangerous fault, was seen as quite clear and caused examiners no second thoughts.

When it was suggested that the system could be amended to ask examiners for marking of ‘general ability’, there was little enthusiasm. The main problem was seen to be how this could be explained to the candidate who had failed the test.

Another suggestion provoked more enthusiasm. It concerned a change of regulations so that if a candidate was marked as having a certain number of minor errors, this would also count as a failure. Some examiners felt that this might be a valuable change and might allow the test to screen-out some of those currently passing despite a generally poor standard of driving.

Instructors were a little less confident in the marking system and the way it is applied, but were nonetheless mostly positive. Their criticisms ranged from claiming that the system is ‘open to interpretation’ on the one hand to feeling that it is ‘inflexible’ on the other. Whilst instructors claimed that they usually know whether a client merits a pass or not, they readily admitted that they cannot predict what exactly will happen during a test. They accepted that some candidates whom they felt were very capable of passing, admitted that they had driven badly and had not deserved to pass. Taking into account the part that they felt luck plays in the system, this deterred them from predicting whether a candidate will pass.

Other possible improvements
Only a small number of other suggestions were made for improvements to the driving test, and these were not generally highly rated by the other participants in those groups where they occurred.

In two groups (one examiners, one instructors) the suggestion was made that all drivers should be re-tested regularly throughout their driving lives. There was fairly widespread agreement that many drivers pick up bad habits, in some cases quickly after their driving test, and more should be done to address this. However, it was recognised that this would have considerable cost and political implications.

In two groups of examiners, suggestions were made that there should be no specific test routes, with examiners free to use their local knowledge around test centres. The advantages cited were that drivers would not be able to ‘learn a route’, and examiners would be free to adjust their route to test aspects of the candidates performance about which they had concerns.
4.2.2 Other issues

Consistency between driving tests

Examiners believed that it was not possible to eliminate variability between driving tests. They recognised that the test centre, the route taken and the time of day could all make a difference to the conditions of the test. Some routes had greater variety of road types, some routes contained specific junctions that caused problems. Similarly it was believed and accepted that the tests starting at 8.40 am and 3.35 pm were generally liable to be made more difficult by the amount of traffic encountered.

It was accepted, therefore, that some tests contained more difficult conditions than others. Examiners generally argued that this variability was not unfair for two reasons: first, the tests all were based on a mix of ‘real life’ driving situations representative of conditions in the candidate’s locality; secondly, they maintained that they made allowances for the specific conditions which applied to each test (their training emphasises the assessment of a fault in the context of prevailing conditions). The Senior Driving Examiners in particular emphasised their ability to be flexible and take conditions into account. However, the younger examiners were more willing to accept that perhaps there was an element of unfairness, or at least randomness in this.

A slightly different type of variation which could be caused by adverse weather conditions was not generally felt to be a major issue: severe weather causes tests to be cancelled, and it was felt that any conditions less severe than this should be within the capabilities of the candidates.

Instructors were less happy about this variability between tests. They felt that chance events sometimes influenced whether a candidate passed or not. These instructors did not generally feel that adequate allowance was made for the different test conditions; they suggested also that the examiners themselves were another source of variability.

There was considerable debate amongst some instructors about specific test centres, routes or examiners thought to present, in their view, unreasonable difficulties. Thus they felt that a candidate’s chances of passing the test depended to a notable degree on luck.

However instructors and examiners shared the opinion that it would not be possible to eliminate variability without totally revolutionising the test procedure. It was inevitable that test conditions in different parts of the country and different types of area would be dissimilar. For the examiners this was not generally a cause of concern, but instructors accepted this a little more reluctantly.

Driving instruction and candidate standards

There was a general feeling among the examiners that the overall standard of candidates taking the test was poor. All accepted that there were some excellent drivers as well as some who were awful, but there was a fairly widespread consensus that the average should be higher.

A wide range of factors were named as contributing to this situation. All groups of examiners participating in the discussions had some criticisms of instructors. In addition, candidates were often felt to go in for tests before they were ready. In some cases this was thought to be because of impatience and an unrealistic opinion of their capabilities, in other cases because of reluctance to keep paying for lessons, whilst a further group were felt to do the test on the principle that, ‘at worst you have had the practice, at best you might pass’. Some of these examiners were critical of the commonly found attitude that people felt that they had a ‘right’ to a licence, rather than feeling that it was something which had to be worked for (by instruction and practice), and earned (by passing the test).

Most, but not all, examiners in the discussion groups believed that people who learned to drive without professional tuition tended to be poorer candidates than those who had lessons. Those learning with their parents were often felt to suffer from out of date or simply incorrect instruction, and the fact that they would come for test in a car without dual controls gave examiners concern about their own safety. On the other hand, candidates who had no opportunities to practice other than driving lessons were likely to have had less total driving experience, so most examiners agreed that the ideal was professional tuition plus some informal practice. Some concluded that a minimum level of professional tuition should be mandatory, others appeared to want private tuition eliminated altogether.

The criticisms made of instructors generally concerned their failure to teach their pupils the whole spectrum of driving skills to an adequate degree. The examiners tended to feel that instructors aimed only to get their pupils through the test, rather than teaching them ‘safe driving for life’. Worse, instructors were generally felt to be poor in achieving their own limited objective. Some examiners spontaneously stated that they could often predict the standard of a candidate by knowing who their instructor was.

Some older and more experienced examiners stated that DSA, and the examiners themselves, had a share of the responsibility for the standard of candidates, since the instructors were felt to work to the perceived test standard. In several groups, examiners expressed some frustration that there was no channel through which they could provide guidance to candidates or instructors about the standards of driving they were seeking.

The instructors, for their part, tended to feel that their professionalism and contribution was undervalued by the public at large and by examiners and the DSA.

Instructors also commented on the problems of teaching pupils who also received instruction (or had previously received instruction) from relatives or friends, with parents particularly picking out for criticism. They often compared UK regulations with other European countries where a specified amount of professional instruction had to be undertaken as a minimum by a learner. Most felt that this should be enforced as the first stage of learning, prior to any practice undertaken elsewhere.

Other aspects of novice driver safety

There was a widespread feeling that young drivers are competent, perhaps skilful drivers, but that the problem was one of attitude. As reported above, some people were felt to ‘put on an act’ to pass their test, after which they switched to their more natural, perhaps aggressive driving style. Whilst some of the measures suggested elsewhere (like a longer test) might be beneficial, this was not
generally seen as an issue which driving instruction or the driving test could really address.

In all groups there was discussion of the law (the Road Traffic (New Drivers) Act 1995, which came into force in 1997) under which recently qualified drivers have their licence revoked and have to re-apply for a provisional licence if they accumulate six or more penalty points within two years. This was largely felt to be a very worthwhile innovation, the only doubts being about the generally low level of law enforcement, and the likely low awareness of the law among its ‘target group’.

Instructors exhibited similar views to the examiners, with perhaps more strongly held feelings about the need to change attitudes among young drivers and potential drivers. They felt in many cases that they had little influence over the likely ‘real’ driving style of their pupils.

When discussion turned to whether newly qualified drivers should display ‘P’ (probationary) plates or green ‘L’ plates to indicate their lack of experience, examiners’ opinions were divided, with few being enthusiastic. Whilst in some cases it was felt that other road users might show more consideration to cars carrying them, others felt that these days many would ignore them or, even worse, rush to avoid getting stuck behind them. There was some unease expressed that they would only be used by drivers lacking confidence.

Some examiners felt more positively about a combination of official ‘P’ plates and restrictions on new drivers, although there was no consensus on which restrictions would be effective. Engine size or power output were the most widely accepted; some supported passenger bans or bans on motorway driving, but others felt that such restrictions were unreasonable. Imposition of a speed limit was felt, after discussion, to be likely to cause dangerous situations. Those who favoured these ideas gave 6 months, 12 months or 2 years as the possible duration of the restricted period. However, there were also several who felt that restrictions would be unenforceable or that they were just unreasonable – on the argument that if drivers had passed their tests, they are entitled to drive without specific restriction.

Instructors also were split on the possible benefits of ‘P’ plates, and on the value of restrictions on newly qualified drivers. Those who were against also mentioned that newly qualified drivers who felt the need for ‘P’ plates reflected badly on their instruction!

4.2.3 Conclusions from examiner and instructor group-discussions

1 Increasing the length of the test was widely called for in the groups because it would (a) increase the range of conditions that could be covered, (b) require sustained concentration from the candidate, (c) improve the examiner’s ability to identify candidates whose on-road behaviour was likely to be a cause for concern, and (d) reduce the chances of a poor candidate passing the test.

2 Reducing the amount of time taken up by the reversing manoeuvres was also seen as a way of increasing the amount of free driving in the test, and achieving some of the above benefits. However, removing the manoeuvres from the on-road test was thought, on balance, to be disadvantageous.

3 Including dual carriageways and national speed limit single carriageway roads in more tests was thought to be very desirable – and was one of the reasons for favouring an increase in test length. There was widespread support for allowing instructors to take learner drivers onto motorways, and for requiring motorway tuition to be certified by instructors either before or after the test.

4 Examiners were generally very satisfied with the test marking system, and did not favour the use of general ratings of ability or safety – mainly on the grounds that it would be difficult to justify failing a candidate on such ‘subjective’ judgements. The idea of failing candidates who exceed a specified number of more minor faults was quite well-liked – partly to reduce the likelihood of having to pass a candidate who was an obviously poor driver, but who happened not to make any serious or dangerous faults during the test. Instructors were somewhat less satisfied with the marking system, but did not suggest any particular improvements.

5 Retesting of drivers at intervals throughout their driving careers, and the abolition of set test routes, were suggested as improvements to the current system, but did not receive much support from other participants.

6 Examiners recognised that there would be inconsistency between tests, associated with traffic conditions, weather, and routes. They accepted this as inevitable and did not see it as a problem. Instructors were more concerned, but did not offer a solution – except in so far as they felt that examiners themselves were one source of inconsistency.

7 Examiners who saw the test as the main guardian of driving standards tended to think that its standard was too low – citing the high accident liability of novice drivers in evidence. Other examiners, and most instructors, felt that the test was only one component in the system. They tended to see the current standard as rather high as evidenced by the poor pass rate.

8 Examiners generally felt that candidates come for test too early, and were very critical of the standard of instruction. They wanted to have a stronger influence on how or what instructors teach their pupils. However, they did think that at least some professional instruction was beneficial. Some felt that all instruction should be professional, but generally a mixture of professional instruction and informal supervision was favoured as a way of maximising pre-test experience. Instructors often felt undervalued by examiners and DSA.

9 Most participants felt that driver attitudes were very important from the point of view of safety, but thought that these could not really be dealt with in the driving test.

10 In principle, the idea of post-test restrictions on novice drivers as a way of improving safety was favoured by most of the examiners and instructors who participated.
5 Factors influencing the pass-rate of the practical driving test

5.1 Introduction

One of the project tasks was to investigate the effects of a number of factors on test pass-rates. Clearly, large differences in pass rate between examiners, types of test route, test centres or different regions of the country would raise questions of fairness, and of whether the intended test standard is being adequately upheld. The same would apply if it were found that test outcome is strongly influenced by the weather, time-of-day, or day of week.

Some variation in pass-rates is unavoidable: it would be impossible to ensure that all test routes posed precisely the same set of demands on candidates, and there are bound to be some differences between examiners. Variations in pass-rates between regions of the country, or between test centres, would pose questions about test standardisation to which there may be no straightforward answer. A high pass rate in a remote, rural part of the country might be symptomatic of easy test routes and undemanding traffic conditions typical of the region. Would this matter, so long as the drivers were competent to drive safely on the roads in the region? From a road safety point of view, the answer would depend on the extent to which drivers restrict themselves to driving in the region, at least during the early post-test period.

Variations in pass-rates between test centres or regions might also reflect differences in the candidates coming forward for test, and do not necessarily imply that the test standard itself is varying. For example, it might be that people from socio-economically deprived areas tend to come forward for test at a relatively low level of competence because they are trying to minimise the cost of learning to drive. If this were so, then low pass-rates in such areas would tend to indicate that the test was successfully helping to uphold standards.

Some types of variation in pass rate would be more obviously undesirable, and potentially easier to deal with. Variations between test routes within a test centre would reduce validity, reliability, and fairness. So, too, would differences between examiners, or effects associated with test time, or weather conditions; though route-differences are not necessarily a particularly important issue, as is discussed further in Section 5.4.

The project included a study in which driving test records, extra information provided by examiners following each test, and information about test routes, were collected and used to explore the effects of the following factors on pass-rates:

- Region of the country (defined in terms of DSA Area).
- Test centre.
- Features of the test route.
- Sex and age of the candidate.
- Sex of examiner.
- Grade of examiner.
- Time of day.
- Day of week.
- Traffic conditions.
- Weather conditions.
- Examiners’ choice of reversing manoeuvres on test.

It should be noted that the findings relate to tests carried out in the last three months of 1997. Since that time, a number of changes have been made to the driving test, and DSA has given considerable attention to both developing an understanding of the factors that contribute towards pass-rate variation between driving test centres and improving quality assurance. Some of that work is mentioned in Section 5.4.

5.2 Data collection

5.2.1 Sample of test centres

A representative sample of test centres was used. The sampling was firstly stratified by the 5 DSA administrative areas, and secondly (within each area), by the ‘size’ of centre, measured by the number of examiners employed. An approximately uniform sampling fraction was used. This meant that the proportion of (say) Wales and Western Area centres in the sample was approximately the same as the proportion of Wales and Western Area centres in the national population of test centres. Similarly, the proportion of (say) ‘small’ test centres in the Wales and Western Area sample was approximately the same as the proportion of small centres in this area as a whole.

Two other factors were taken into account when drawing the test sample from the Scotland Area. Test centres in Scotland included a number of temporary ‘Islands’ centres on some of the larger Islands. The ratio of Islands to Mainland centres in the Scotland sample (of 6 centres) was set to reflect the ratio of Islands to Mainland centres in Scotland. This resulted in having two Islands centres in the sample. Secondly, for those centres on the Scottish mainland that had ‘Outstations’ (where there was no permanently stationed Senior Examiner) the tests conducted at these outstations were also included in the sample for the main centre.

The final sample comprised 7 centres from the London and South East Area, 10 from the Midlands and Eastern Area, 8 from Wales and Western area, 11 from the Northern Area and 6 from Scotland. It should be noted that centres in London and the South East Area tend to have more examiners (and test routes), and conducted more tests than in other Areas, although each Area does have ‘larger’ centres within the major towns and cities.

5.2.2 Examiner questionnaire

In addition to the normal driving test form (DL.25) on which examiners record the faults made during a driving test, examiners were asked to complete a separate questionnaire. This provided the following information:

- The driving test centre code, examiner number, date and time of test – forming the unique ‘test’ identifier,
- Age of the candidate.
- Number of tests taken (although this was completed only if the driver passed).
Two indicators of the traffic conditions on test:

i. How difficult were general traffic conditions in terms of passing or failing the test? (Scale points: 1 - very easy, 2 - easy, 3 - average, 4 - difficult, and 5 - very difficult.) The assessment was made in comparison with average traffic conditions for all the tests (all routes and all times) conducted at the centre.

ii. What proportion of the time was spent in ‘free’ driving (as opposed to being in slow moving traffic, waiting in stationary traffic, or conducting the manoeuvres)? (Categories: less than one-quarter; between one-quarter and a half; between a half and three-quarters; more than three-quarters).

Weather conditions: sunny day/dry roads; overcast day/dry roads; not raining but wet roads; light rain/drizzle; medium rain; heavy rain; other conditions.

5.2.3 Route characteristics
At all the test centres in the sample, each test route was driven by an examiner who completed two forms that summarised the features of the route. The information they provided included the following:

- Route length in miles.
- Percentage of the route falling into the following four categories: ‘country/rural’, ‘30 mph or less’, ’40 mph’, ‘50 mph or more’. Note: with very few exceptions, these categories were treated as mutually exclusive, so that ‘country/rural’ equated to 60/70 mph sections.
- Whether or not the following carriageway feature existed on the route: single carriageway section, dual carriageway section, single with 3 or more lanes, dual with 3 or more lanes in the same direction, single lane one-way streets, or multi-lane one-way streets.
- Whether or not the following speed limits were encountered on the route: 20, 30, 40, 50, 60 and 70 mph.
- The number of particular ‘features’ on the route. Features included traffic lights, pedestrian crossings – light controlled and uncontrolled, roundabouts (mini, single-lane and large), bends, hill crests, box junctions, school entrances and crossing patrols, and level crossings/moving bridges.
- The number of manoeuvres executed on the route, recorded in eight categories – ‘going ahead’, ‘turning right’, ‘turning left’, and ‘use of slip roads’ – each at either signal controlled or non-signal controlled junctions.

This information was available for 643 routes, which represents all but one of the routes used for the sample of driving tests included in the study.

5.2.4 The combined data
22,476 matched records (DL25 form plus examiner questionnaire) were available for analysis. The analyses of the effects of route characteristics used the same basic file but with 272 tests missing – those corresponding to the single route for which the data were missing.

5.3 Results
Sections 5.3.1 and 5.3.2 which follow, present analyses of pass-rates as a function of age and sex, day of week and time-of-day, and by test centre and examiner, based on the data in the combined file described above. Section 5.3.3 considers aspects of the type of test route used and the driving conditions encountered.

5.3.1 Candidate age and sex, and time-of-test effects
Candidate age and sex effects
Logistic regression was used to explore the effects of age and sex on pass-rates – and Figure 2 illustrates the result. At all ages, male candidates had higher pass-rates than females. Pass-rates declined with age, the decline for females being steeper than that for males. All these effects were statistically significant. These sex and age effects are large and consistent and they need to be taken into account when assessing pass-rates as a function of some of the other variables in the analysis.

![Figure 2 Effect of age and sex on pass rates](image)

Day-of-week
Table 1 shows the total number of tests carried out during the 12 week survey period by day of week (Monday to Saturday) and (in column 3) the average pass rate on each day averaged across the survey period.

<table>
<thead>
<tr>
<th>Day of week</th>
<th>Number of tests</th>
<th>Pass rate%</th>
<th>Estimated pass rate of a 17 year old male %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>4,659</td>
<td>47.8</td>
<td>55.9</td>
</tr>
<tr>
<td>Tuesday</td>
<td>4,750</td>
<td>47.1</td>
<td>55.1</td>
</tr>
<tr>
<td>Wednesday</td>
<td>4,077</td>
<td>47.1</td>
<td>54.9</td>
</tr>
<tr>
<td>Thursday</td>
<td>4,303</td>
<td>48.1</td>
<td>55.7</td>
</tr>
<tr>
<td>Friday</td>
<td>4,301</td>
<td>45.3</td>
<td>53.3</td>
</tr>
<tr>
<td>Saturday</td>
<td>386</td>
<td>47.9</td>
<td>53.8</td>
</tr>
<tr>
<td>Average over days</td>
<td>22,476</td>
<td>47.2</td>
<td>54.8</td>
</tr>
</tbody>
</table>
into account by including age and sex in the statistical model ($\chi^2 = 6.64, p = 0.01$). However, when individual examiner effects (see Section 5.3.2) are added to the model, the Friday effect ceases to be significant at the 5% level, although in the sample, Friday’s pass-rates are still lower that the overall mean.

The final column in Table 1 shows the pass-rates of 17 year old male drivers estimated from a logistic regression model which includes age, sex and examiner effects as well as the day of week effect.

**Time-of-day**

Driving tests are carried out in well defined ‘time slots’ each day. With a very few exceptions the starting times of tests in the present sample were: 8.40, 9.30, 10.30, 11.20, 12.10, 13.55, 14.45 and 15.35. Table 2 (columns 1-3) show the total number of tests in the sample carried out at these times, and the average pass rate in each time slot.

**Table 2 Number of tests and pass-rates by time slot**

<table>
<thead>
<tr>
<th>Time slot</th>
<th>Number of tests</th>
<th>Pass rate</th>
<th>Percentage of female candidates</th>
<th>Average age of female candidates (yrs)</th>
<th>Pass rate of 17 year old male (predicted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.40</td>
<td>2,143</td>
<td>49.0</td>
<td>53.1</td>
<td>23.8</td>
<td>56.5</td>
</tr>
<tr>
<td>9.30</td>
<td>3,245</td>
<td>47.5</td>
<td>58.4</td>
<td>25.3</td>
<td>55.5</td>
</tr>
<tr>
<td>10.30</td>
<td>3,395</td>
<td>45.5</td>
<td>62.9</td>
<td>25.5</td>
<td>53.9</td>
</tr>
<tr>
<td>11.20</td>
<td>3,279</td>
<td>45.4</td>
<td>60.1</td>
<td>25.1</td>
<td>53.9</td>
</tr>
<tr>
<td>12.10</td>
<td>2,979</td>
<td>48.0</td>
<td>55.6</td>
<td>24.5</td>
<td>55.7</td>
</tr>
<tr>
<td>13.55</td>
<td>3,044</td>
<td>47.0</td>
<td>54.7</td>
<td>24.9</td>
<td>54.7</td>
</tr>
<tr>
<td>14.45</td>
<td>2,157</td>
<td>46.6</td>
<td>52.7</td>
<td>23.8</td>
<td>53.8</td>
</tr>
<tr>
<td>15.35</td>
<td>2,234</td>
<td>49.1</td>
<td>49.9</td>
<td>23.5</td>
<td>56.3</td>
</tr>
<tr>
<td>All times</td>
<td>22,476</td>
<td>47.3</td>
<td>56.5</td>
<td>24.7</td>
<td>55.0</td>
</tr>
</tbody>
</table>

It can be seen (Table 2, column 3) that there are noticeable differences in pass rate in the different time slots. In particular, pass-rates in the tests beginning at 10.30 and 11.20 are appreciably lower than the pass rate averaged across time slots (47.3%) – and these differences turn out to be statistically significant at better than the 5% level. However, it is necessary to check whether these raw pass rate variations can be accounted for by the differences in the age and sex distributions between the particular time slots. Columns 4 and 5 of Table 2 shows the proportion of female candidates in each time slot and the average age of the female candidates (there is no significant variation in the age of the male candidates from time slot to time slot). It can be seen that there is a systematic pattern in both of these variables, such that in the time slots beginning 10.30 and 11.20, (a) there tended to be a higher proportion of women candidates and (b) the women tested at these times tended on average to be slightly older than those tested either in the first time slot or those tested at the end of the day. Both of these differences will have had the effect of depressing the pass-rates during the time slots starting at 10.30 and 11.20.

Pass-rates were adjusted for these differences by including age and sex in the statistical model. After this adjustment, the time slots beginning at 10.30 and 11.20 still produce lower pass-rates, but these rates are not now significantly different from the overall rate (the $\chi^2$ values fall from 4.44, ($p < 0.05$) to 1.05, ($p = 0.31, n.s.$) for the 10.30 slot and from 4.99, ($p < 0.05$) to 2.51, ($p = 0.11, n.s.$) for the 11.20 slot). Adding individual examiner effects to the statistical model does not change this result.

The final column in Table 2 (column 6) shows the pass-rates of 17 year old male drivers predicted from a statistical model which includes age, sex, examiner effects and time slot. It is reasonable to conclude therefore that the differences in pass-rates shown in column 3 of the Table 2 are principally due to age and sex effects. Although predicted pass-rates for 17 year old male candidates in the first two slots and the last one remain slightly higher than in the remainder, these differences are not statistically significant.

To check whether there is a day of week or time of day effect, the pass-rates during the test slots starting at 08.40 and 09.30 on Monday mornings, and those starting at 14.45 and 15.35 on Friday afternoons, and tests conducted at any time on Saturdays were compared with the overall pass rate. Using a statistical model to adjust for the effects of age, sex and individual examiners showed that the Monday morning pass rate was a statistically significant 3 percentage points higher than the overall pass rate. The slightly lower pass-rates on Saturdays and in the last two slots on Fridays (about 1.2 percentage points lower) were not significantly different from the overall value. In view of the fact that the pass-rates in the first two tests on any day of the week tend to be higher than those in the middle of the day, the apparent Monday morning effect may be a general morning effect rather than one specifically associated with Mondays.

**5.3.2 Examiner, centre and regional effects**

**Introduction**

The analyses presented in the next three sections examine whether there are variations in pass rate arising from differences in examiners, test centres and geographic areas.

**Examiners**

Test results from 258 examiners drawn from 5 areas of the country and 42 test centres were available for analysis. The number of tests carried out by any one examiner in the sample ranged from 2 to 307. To maintain accuracy in the estimation of pass-rates some of the results presented in this section will be limited to the 102 examiners who carried out over 100 tests. The distribution of the pass-rates of these 102 examiners will also be compared with DSA data giving the distribution of pass-rates for all examiners in the country who carried out more than 500 tests during the whole of 1997.

First, however, the effects of examiner sex and grade will be evaluated using pass-rates derived from all examiners in the sample (except one for whom information about the sex and grade was missing). The results in the
following two sections therefore relate to 257 examiners and 22,467 tests.

Examiner gender
Of the 257 examiners whose gender was given, 228 (89%) were male and 29 (11%) were female. Table 3 shows the test pass-rates by sex of the examiner and sex of the candidate. The overall pass rate for female examiners of 46.9% (the final column of Table 3), was not significantly different from the 47.1% pass rate for the male examiners ($\chi^2 = 0.013, p = 0.91, \text{n.s.}$). Moreover there was no observable interaction between examiner sex and candidate sex ($\chi^2 = 0.10, p = 0.75, \text{n.s.}$) – in other words, whatever the candidate’s sex, likelihood of passing is not dependent on the sex of the examiner.

Table 3 Pass-rates (and number of tests) by sex of examiner and candidate

<table>
<thead>
<tr>
<th>Sex of examiner</th>
<th>Male candidates</th>
<th>Female candidates</th>
<th>All tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>52.0% (8,627)</td>
<td>43.4% (11,147)</td>
<td>47.1% (19,768)</td>
</tr>
<tr>
<td>Female</td>
<td>52.2% (1,147)</td>
<td>42.9% (1,552)</td>
<td>46.9% (2,699)</td>
</tr>
</tbody>
</table>

Examiner grade
Examiners were grouped into four ‘grades’ – Senior Driving Examiners (SDE), Driving Examiners (DE), Re-employed retired examiners (FP), and Contract Examiners (C). Of the 257 examiners for whom the information was available, 34 were Senior Driving Examiners (of whom 3 were women), 171 were Driving Examiners (of whom 17 were women), 12 fell into the re-employed category (of whom 1 was a woman) and 40 were Contract Examiners (of whom 8 were women). The pass-rates of these grades by sex of the candidates are shown in Table 4.

Table 4 Pass-rates (and number of tests) by grade of examiner and sex of candidate

<table>
<thead>
<tr>
<th>Grade of examiner</th>
<th>Male candidates</th>
<th>Female candidates</th>
<th>All tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDE</td>
<td>51.3% (1,268)</td>
<td>44.3% (1,801)</td>
<td>47.2% (3,069)</td>
</tr>
<tr>
<td>DE</td>
<td>51.8% (7,003)</td>
<td>41.8% (8,955)</td>
<td>46.2% (15,958)</td>
</tr>
<tr>
<td>FP</td>
<td>57.1% (56)</td>
<td>48.3% (89)</td>
<td>51.7% (145)</td>
</tr>
<tr>
<td>C</td>
<td>53.5% (1,447)</td>
<td>49.4% (1,848)</td>
<td>51.2% (3,295)</td>
</tr>
</tbody>
</table>

For overall pass rate, the only statistically significant effect was that the pass rate of Contract Examiners was higher than that of the Driving Examiner grade by about 5 percentage points ($\chi^2 = 27.4, p < 0.001$). This conclusion remains essentially unaltered when any differences in the age and sex structure of the groups of candidates tested by the various grades is allowed for by adding age and sex to the statistical model.

Table 4 (columns 2 and 3) can be examined to investigate the interactions between examiner grade and the sex of the candidate. Compared to the Driving Examiner grade, Senior Driving Examiners have a slightly lower pass rate for male candidates and a slightly higher pass rate for female candidates; these differences are not however statistically significant at the 5% level ($\chi^2 = 2.2, p = 0.14, \text{n.s.}$). Contract Examiners also show a similar interaction with a pass rate for male candidates which is somewhat higher (by 1.7 percentage points) than those of the Driving Examiner grade and a pass rate for female candidates which is 7.6 percentage points higher. This effect is statistically significant ($\chi^2 = 9.7, p < 0.01$).

Variations between individual examiners: the study sample
This section presents the distribution of pass-rates between individual examiners of both sexes and all grades, estimated from the sample of tests included in the present study. The accuracy with which the pass rate of an examiner can be determined will depend on the number of tests available in the sample carried out by any particular examiner a number that varied between 2 and 307 in this study. To achieve reasonable levels of accuracy only those 102 examiners for whom data were available on at least 100 tests were used in the present analysis.

100 tests will yield a 95% confidence interval on an estimate of examiner pass rate of about ± 10 percentage points; the corresponding confidence interval for 300 tests (the largest sample in the data from one examiner) would be ± 6 percentage points. These figures represent the variation in pass-rates arising solely from the inherent variability of the binary data; there will of course be other sources of variation, including the fact that the sample is limited to a time window of a few weeks or months (100 tests would be undertaken over a period of at least three weeks). A more satisfactory estimate of pass rate would require sampling over a full year; however, as will be seen in the following section, the sample distribution presented in this section gives a very realistic picture of the distribution of pass-rates among examiners.

Figure 3 is a histogram showing the distribution of pass-rates from the 102 examiners in the present sample who carried out at least 100 tests in the study period. It will be

Figure 3 Distribution of pass rates for those examiners with 10 or more tests in the sample
seen that there is a considerable spread of pass-rates; the distribution shown in the figure has a mean of 46.7% and a standard deviation of 8.1%.

The distribution shown in Figure 3 includes the variation arising from the relatively poor accuracy with which individual pass-rates can be estimated from samples of tests carried out by each examiner. If this source of error was of a magnitude corresponding to a sample size of 177 tests (the average for this group of examiners), the standard deviation of the underlying examiner pass rate distribution would fall to 7.2%. This latter figure means that 68% of examiners will have a pass rate lying between 39.5% and 53.9%, and 95% of examiners will have a pass rate lying between 32.6% and 60.8%. Note, however, that much of this variation may stem from differences between the test centres at which these examiners worked. This is discussed further below.

As in the case of areas and test centres, the examiner pass rate data was checked to see whether variations the age and sex distributions of candidates from examiner to examiner made any difference to the pass-rates. No effect was found.

Variations between examiners – DSA data for 1977

To check that the results obtained from the sample were not atypical of a full year’s data, DSA provided pass-rates for all examiners for the whole of 1997. Again, to avoid excessive variability due to poor estimation of pass-rates for individual examiners, only examiners who had carried out at least 500 tests in the year were included in the analysis of these data.

Figure 4 shows the distribution obtained from 988 examiners who had carried out an average of 1084 tests each during 1997. The mean of the distribution is 46.6% and the standard deviation 7.6%. Making a small correction as above for statistical variability in the measurement of pass rate, the underlying standard deviation of the distribution would be about 7.4%. Thus the mean and standard deviation of the distribution shown in Figure 4 from a full year’s test data are very similar to the values estimated from the sample as presented in the previous section.

To demonstrate that those examiners who had a high pass rate in the study sample also had a high pass rate for the year as a whole – again based on those examiners in the sample who had conducted 100 or more tests – Figure 5 shows their pass-rates in the sample period plotted against the pass rate for the year as a whole. Clearly, examiners’ pass-rates in the sample period correspond closely to their pass-rates in the year as a whole; the correlation coefficient is 0.86.

Variability of examiner pass-rates within test centre

Examiners generally work within a particular test centre or in a few geographically close test centres. In the data available to this study, each examiner’s pass rate data was associated with one specific test centre, and Figure 3 shows the variation of examiner pass-rates regardless of which test centre they worked in whilst Figure 4 shows that this is representative of the variation in examiner pass-rates over a full year.

If the variation between the average pass-rates at centres (shown in Table 5 for the present sample) were determined wholly by factors associated with centre (eg by the quality of the candidates submitting themselves for test), and not at all due to differences in examiners between centres then the variation in examiner pass-rates shown in Figures 3 and 4 would be a considerable over-estimate because it would include the source of variation arising from centre differences. In order to present the variation in examiner pass-rates with this source of variation removed, Figure 6 shows the distribution of the deviations of individual examiner’s pass-rates from that of the centre in which the tests were carried out – once again using only those examiners who had conducted 100 tests or more.

Since the mean for each test centre has been subtracted from the examiner pass-rates, the distribution shown in Figure 6 has a mean of zero. The standard deviation is 5.4%, which would fall to about 4% if an allowance was made for the variability of the basic measurement of pass rate (based again on a average of 177 tests per examiner). Thus, 68% of examiners will have a pass rate lying between 4 percentage points below and 4 percentage points above the centre average, and 95% of examiners will have a pass rate lying between 8 percentage points below and 8 percentage points above the test centre averages.
Table 5 Number of tests and pass-rates by test centre

<table>
<thead>
<tr>
<th>Test centre</th>
<th>Area</th>
<th>Number of tests</th>
<th>Average pass rate %</th>
<th>Predicted pass rate for a 17 year old male driver %</th>
<th>Statistical significance of difference from average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lerwick (Shetland)</td>
<td>SC</td>
<td>58</td>
<td>70.7</td>
<td>76.1</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Carmarthen</td>
<td>W&amp;W</td>
<td>143</td>
<td>67.8</td>
<td>72.7</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Barnstable</td>
<td>W&amp;W</td>
<td>306</td>
<td>61.8</td>
<td>67.4</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Perth</td>
<td>SC</td>
<td>217</td>
<td>61.7</td>
<td>68.1</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Aberdeen (Balgownie Rd.)</td>
<td>SC</td>
<td>383</td>
<td>59.8</td>
<td>65.9</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Skipton</td>
<td>N</td>
<td>233</td>
<td>57.1</td>
<td>63.0</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Inverness</td>
<td>SC</td>
<td>214</td>
<td>56.1</td>
<td>61.7</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Cardiff (Cathays Terrace)</td>
<td>W&amp;W</td>
<td>522</td>
<td>54.8</td>
<td>61.5</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Kendal</td>
<td>N</td>
<td>379</td>
<td>54.1</td>
<td>60.0</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Eastbourne</td>
<td>LSE</td>
<td>617</td>
<td>53.5</td>
<td>59.4</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Manchester (Cheetham Hill)</td>
<td>N</td>
<td>763</td>
<td>52.0</td>
<td>59.1</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Swindon</td>
<td>W&amp;W</td>
<td>665</td>
<td>51.4</td>
<td>57.5</td>
<td>ns</td>
</tr>
<tr>
<td>Bristol (Brislington)</td>
<td>W&amp;W</td>
<td>735</td>
<td>51.3</td>
<td>57.0</td>
<td>ns</td>
</tr>
<tr>
<td>Newtown</td>
<td>WSW</td>
<td>82</td>
<td>51.2</td>
<td>56.7</td>
<td>ns</td>
</tr>
<tr>
<td>Loughton</td>
<td>LSE</td>
<td>998</td>
<td>51.2</td>
<td>57.3</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Telford</td>
<td>M&amp;E</td>
<td>648</td>
<td>49.8</td>
<td>56.2</td>
<td>ns</td>
</tr>
<tr>
<td>Plymouth</td>
<td>W&amp;W</td>
<td>822</td>
<td>49.3</td>
<td>54.8</td>
<td>ns</td>
</tr>
<tr>
<td>Glasgow (Shawlands)</td>
<td>SC</td>
<td>1025</td>
<td>48.2</td>
<td>56.1</td>
<td>ns</td>
</tr>
<tr>
<td>Birmingham (Kingstanding)</td>
<td>M&amp;E</td>
<td>371</td>
<td>46.6</td>
<td>53.1</td>
<td>ns</td>
</tr>
<tr>
<td>Norwich (Jupiter Rd.)</td>
<td>M&amp;E</td>
<td>1167</td>
<td>46.4</td>
<td>52.1</td>
<td>ns</td>
</tr>
<tr>
<td>Sheffield</td>
<td>N</td>
<td>638</td>
<td>46.1</td>
<td>53.4</td>
<td>ns</td>
</tr>
<tr>
<td>Wealdstone</td>
<td>LSE</td>
<td>878</td>
<td>46.0</td>
<td>52.9</td>
<td>ns</td>
</tr>
<tr>
<td>Pinner</td>
<td>LSE</td>
<td>956</td>
<td>45.8</td>
<td>52.2</td>
<td>ns</td>
</tr>
<tr>
<td>Hornchurch</td>
<td>LSE</td>
<td>801</td>
<td>45.2</td>
<td>51.3</td>
<td>ns</td>
</tr>
<tr>
<td>Crawley</td>
<td>LSE</td>
<td>813</td>
<td>45.1</td>
<td>50.0</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Liverpool (Norris Green)</td>
<td>N</td>
<td>1083</td>
<td>44.8</td>
<td>53.1</td>
<td>ns</td>
</tr>
<tr>
<td>Nottingham (West Bridgeford)</td>
<td>M&amp;E</td>
<td>409</td>
<td>44.7</td>
<td>50.9</td>
<td>ns</td>
</tr>
<tr>
<td>Bedford</td>
<td>M&amp;E</td>
<td>930</td>
<td>44.4</td>
<td>49.9</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Hereford</td>
<td>M&amp;E</td>
<td>397</td>
<td>43.6</td>
<td>49.4</td>
<td>ns</td>
</tr>
<tr>
<td>Oxford (Cowley)</td>
<td>M&amp;E</td>
<td>314</td>
<td>43.3</td>
<td>49.2</td>
<td>ns</td>
</tr>
<tr>
<td>Chester</td>
<td>N</td>
<td>181</td>
<td>43.1</td>
<td>49.2</td>
<td>ns</td>
</tr>
<tr>
<td>Scarborough</td>
<td>N</td>
<td>238</td>
<td>42.9</td>
<td>49.4</td>
<td>ns</td>
</tr>
<tr>
<td>Carlisle</td>
<td>N</td>
<td>682</td>
<td>42.8</td>
<td>48.6</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Sunderland</td>
<td>N</td>
<td>435</td>
<td>42.8</td>
<td>49.7</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Newbury</td>
<td>W&amp;W</td>
<td>423</td>
<td>41.1</td>
<td>46.0</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>York</td>
<td>N</td>
<td>884</td>
<td>41.1</td>
<td>47.1</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Brentwood</td>
<td>LSE</td>
<td>368</td>
<td>41.0</td>
<td>46.4</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Bishop Auckland</td>
<td>N</td>
<td>490</td>
<td>39.4</td>
<td>44.9</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Leicester (Welford Rd.)</td>
<td>M&amp;E</td>
<td>397</td>
<td>38.0</td>
<td>44.8</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Rugby</td>
<td>M&amp;E</td>
<td>425</td>
<td>36.2</td>
<td>41.6</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Lincoln</td>
<td>M&amp;E</td>
<td>377</td>
<td>30.8</td>
<td>35.9</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Isle of Skye</td>
<td>SC</td>
<td>9</td>
<td>22.3</td>
<td>29.4</td>
<td>ns</td>
</tr>
</tbody>
</table>

All centres                      |      | 22,476          | 47.7                | 54.0                                                 |                                                     |
the age and sex structure of the samples of candidates in the pass rate table – indicating that differences in has made only minor changes to the relative position of the can be seen that the addition of age and sex to the model different from the overall average (ns = not significant). It should be pointed out here that with multiple comparisons of the sort being made in Table 5, we would expect one in 20 of the differences found to be significant at the 5% (p≤0.05) level to be significant by chance.

Area of country
Candidates taking the test were sampled from test centres from each of the five DSA administrative areas of Great Britain. Table 6 shows the pass-rates by area in descending order.

### Table 6 Number of tests and pass-rates by area of the UK

<table>
<thead>
<tr>
<th>Area</th>
<th>Number of tests</th>
<th>Average pass rate</th>
<th>Estimated pass rate for a 17 year old male driver</th>
<th>For column 4 statistical significance of difference from average over all areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midlands and Eastern</td>
<td>5,435</td>
<td>43.5</td>
<td>49.4</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>North of England</td>
<td>6,006</td>
<td>45.4</td>
<td>52.2</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>London and the SE</td>
<td>5,431</td>
<td>47.2</td>
<td>53.2</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Wales and the SW</td>
<td>3,698</td>
<td>51.7</td>
<td>57.4</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Scotland</td>
<td>1,906</td>
<td>53.5</td>
<td>60.6</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>All areas</td>
<td>22,476</td>
<td>48.3</td>
<td>54.6</td>
<td></td>
</tr>
</tbody>
</table>

Table 6 (column 3) shows that the average pass-rates (uncorrected for candidate age and sex differences) vary from one area of the country to another. With the exception of London and the SE (which is in any case close to the average pass rate of 48.3%) the uncorrected pass-rates (column 3) in the other four areas are all significantly different from the average. Column 4 shows the pass-rates for 17 year old males predicted from a logistic regression model which includes age and sex, and the final column (5) indicates the statistical significance of the differences between these ‘corrected’ values and the corresponding pass rate averaged over areas.

It will be seen that the correction for age and sex has slightly increased the difference between the London pass rate and the average, such that it is now significant at the 5% level.

It is of course to be expected that pass-rates will vary from area to area, from centre to centre and from examiner to examiner for all kinds of reasons. Table 6 shows that the test centres in the Midlands and Eastern Area have the lowest pass-rates followed by those in the North of England. The highest pass-rates are to be found in Scotland, where the pass-rates are slightly higher than...
those in Wales and SW England. London and the South East has a pass rate not far from the overall average. The variations between areas (and centres) may result from a variety of factors associated with the area or the centre – including the competence of candidates offering themselves for test\(^1\). In this sample the area pass-rates are including the competence of candidates offering a variety of factors associated with the area or the centre – variations between areas (and centres) may result from a variety of factors associated with the area or the centre.

5.3.3 The effect of individual route characteristics and driving conditions on pass-rates

Introduction

The method used to determine whether a particular route, traffic or weather feature encountered on test was a significant determinant of pass rate, was to fit a ‘baseline’ logistic model of pass rate which included age and sex of the candidate and examiner effects, and then to add the feature of interest to this model and to test whether or not its addition significantly improved the model’s fit – ie its ability to predict pass-rates.

Because of the hierarchical structure of the data available in this study it is difficult to be certain which way the causal links between examiner pass-rates and route/traffic effects operate. The approach used therefore in the present study was always to include the age, sex and examiner effects in the statistical model along with the route variables of interest. The strengths of the various effects (and their implicit interactions) are determined by the maximum likelihood procedure used by the statistical modelling algorithm. This process determines the values of the coefficients in the model to be those which maximise the probability that the observed data would actually occur. Also, because a large number of route and traffic variables were collected (over 30) one or two might be expected to appear significant at the 5% level by chance. For this reason, a 1% level of significance is more appropriate at this part of the analysis.

The number of features on the route

The actual number of the following static features which occurred on the test route were recorded: traffic lights, pedestrian crossings not controlled by lights, light-controlled pedestrian crossings (pelicans), location of school crossings, severe bends, school entrances, level crossing/moving canal bridges, narrow hill crests, box junctions, and roundabouts (mini, single-lane and multi-lane).

Once age, sex and examiner effects had been included in the logistic model, only light-controlled pedestrian crossings, single-lane (not mini) roundabouts and multi-lane roundabouts were significantly predictive of pass rate. The direction of all these effects was that the greater the severity of the feature of interest to this model and to test whether or not its addition significantly improved the model’s fit – ie its ability to predict pass-rates.

The numbers of light controlled crossings and roundabouts (excluding mini roundabouts) can be conveniently replaced by a combined variable F (the sum of the numbers of pelicans and roundabouts), the 5 to 95 percentile range of which is from 1 to 17. The predicted pass rate of a 17 year old male candidate on a route with only 1 pelican crossing or roundabout on it will be 60.5%; at the other extreme, on a route with 17 crossings or roundabouts the same 17 year old male would have a pass rate of 47.5%; ie 13 percentage points lower\(^2\).

The question arises of the extent to which between-route differences in number of light-controlled crossings and roundabouts may explain between-examiner differences in pass-rates. Examiners use several routes during the course of their work, so the impact of these route differences on examiner pass-rates could not be as large as the 13 percentage points mentioned above. In fact, using mean values of F for the 100 or so examiners in the present sample for whom route information was available and who had carried out more than 100 tests, and adjusting their pass-rates for these differences, yielded a distribution of ‘corrections’ which had a standard deviation of 3.3 percentage points in pass rate. The ‘correction’ estimated in this way proved to be virtually independent of the examiner basic pass rate, so that adjusting examiner pass-rates for this route effect did not reduce the variation between examiners shown in Figures 3 and 6 above.

The presence or absence of particular road-types on the route

Routes were also classified according to whether they had sections corresponding to the following road types: single carriageways, dual carriageways, single carriageways with three (or more) lanes, dual carriageways with three (or more) lanes in the same direction, one way streets (single-lane) and one way streets (multi-lane).

A logistic model which included age, sex and examiner effects suggested that test routes which included a section of dual carriageway with three or more lanes in the same direction had significantly lower pass-rates compared with the remainder (\(\chi^2 = 18.3, p < 0.001\)). The 11% of routes with such a feature on them accounted for 47.5% of the remainder (\(\chi^2 = 5.9, p < 0.001\)). The size of the effect was such that those routes with a section of 3-lane dual carriageway produced a pass rate which was 5.9 percentage points lower than the remainder. Such an effect could possibly arise from the challenging task of joining or leaving the dual carriageway from a slip road, and the number of joining or exiting slip road manoeuvres at dual carriageways was actually recorded as part of the route information. However, neither the number of joining movements nor the number of exiting movements at non-signalised junctions with dual carriageways proved to be significant predictors of pass-rates. This result suggests that joining or leaving a dual carriageway is not a manoeuvre which generates serious or dangerous faults on test, and leaves open the question of

\(^1\) Subsequent work by DSA has provided further evidence here – see Section 5.4

\(^2\) Pass rate is calculated as \(P = \frac{1}{1 + e^{-Z}}\), and the equation predicting the Z score of a 17 year old male is: \(Z = 0.46 - 0.033F\). Thus the values of Z for \(F=1\) and \(17\) are respectively \((0.46 - 0.033) = 0.427\) and \((0.46 - 0.561) = -0.101\). Therefore, for \(F = 1\), \(P = \frac{1}{1 + e^{0.427}} = 0.605\) (60.5%), and for \(F = 17\), \(P = \frac{1}{1 + e^{-0.101}} = 0.475\) (47.5%).
why these particular dual carriageways are associated with low pass-rates. The finding implies that 3-lane carriageways are associated with some other factor that makes the test ‘difficult’ – perhaps the level of traffic flow on these roads, or on test routes in areas that have these roads, or the difficulty of the lane selection and lane changing task.

To explore the effect on pass-rates of dual carriageways in general, a logistic model was fitted using a variable which identified whether routes included a dual carriageway section of any type (61% of routes did). The resulting model showed that the pass-rates on routes which included some dual carriageways (of whatever type) were only 0.5 percentage points lower than the remainder – an effect which was not statistically significant ($\chi^2 = 0.2, p = 0.65, \text{n.s.}$). A similar result was obtained in relation to the presence or absence of a single-lane one-way street. The proportion of routes which incorporated a section of single-lane one-way street was just over 18%, accounting for 16% of tests. The size of the effect was such that routes which contained such a section had a pass rate which was 2.8 percentage points lower than the remainder ($\chi^2 = 6.4, p = 0.01$). The corresponding effect for routes which included sections of one-way street of any kind (single or multi-lane – averaging 36% of routes) was that such routes had a pass rate which was only 1.2 percentage points lower than the remainder – again a non-significant difference ($\chi^2 = 1.6, p = 0.21, \text{n.s.}$).

**Speed limits**

Two types of data were collected which related to speed limits on the route. The first asked for estimates of the proportion of four speed limit categories on the route. The categories were ‘30 mph or less’, ‘40 mph’, ‘50 mph or more’ and ‘country/rural roads’. As has been indicated earlier respondents generally treated these four categories as non-overlapping (so that the percentages summed to 100). The alternative method of associating speed limits with routes was to record whether the route contained a section with a 20 mph, 30 mph, 40 mph, 50 mph, 60 mph or 70 mph speed limit on it or not – a series of binary variables.

A logistic regression using these variables alongside age, sex and examiner effects suggested that none of these speed-limit related variables could be regarded as significant or reliable predictors of pass rate.

**Turns at signalised or non-signalised junctions**

The route information also contained a measure indicating how many different types of manoeuvre were required to navigate the route. However, none of the data relating to manoeuvres on routes (the number of ahead movements or the number of right or left turns, whether at signalised junctions, non-signalised junctions or dual carriageways) proved to be significant as predictors of pass rate.

**Traffic conditions**

Examiners completing the post-test questionnaire were asked to assess traffic conditions on a five point scale: very easy, easy, average, difficult, very difficult. They were asked to estimate this relative to the average conditions over all the tests (all routes and all times) conducted at that test centre. The relationship between these assessments and pass rate were not consistent nor were they statistically significant at the 1% level; if anything, pass-rates were somewhat lower than the overall average when the traffic conditions were assessed as easy or average. Accordingly, it seems reasonable to assume that traffic conditions (at least as assessed by examiners, and relative to average conditions for their test centre routes) are not an important determinant of pass rate.

*‘Free driving’ time*

Examiners also estimated on a four point scale what proportion of time during the test had been spent in ‘free’ driving as opposed to driving in slow moving traffic, waiting in stationary traffic or conducting manoeuvres. The four point scale was: 1 - less than one-quarter; 2 - between one-quarter and a half; 3 - between a half and three-quarters; 4 - more than three-quarters.

When this variable was included in a statistical model for pass rate which also included age, sex and examiner effects, the resulting coefficients of the four ‘free driving’ categories varied systematically: lower proportions of time spent free driving on test resulted in lower pass-rates and higher proportions of time spent free driving in higher pass-rates. Actually, only the fourth of these categories (more than three-quarters of the time spent free driving) proved to be statistically significant as an individual category ($\chi^2 = 21.6, p < 0.001$), but the trend for pass rate to increase with free driving time was consistent across all four categories. Because of this, the free driving time scale parameter was included in a model as a single continuous variable $T$ (range 1 to 4); the coefficient of 0.16 for $T$ was significant ($\chi^2 = 25.9, p < 0.001$). This means that the estimated pass rate of a 17 year old male candidate driving when less than one-quarter of the test is ‘free’ driving would be 46.6% whilst the same candidate driving under conditions in which over three-quarters of the test was ‘free’ driving would have a pass rate of 58.5% – a difference of 11.9 percentage points.

Once again, the points made about the impact of the number of features on the route (the variable $F$ in Section 5.3.3) on examiner pass-rates is equally true of the driving time variable. As far as examiners are concerned the effect of differences in driving time will to some extent average-out across tests, though there will remain some area effects due to variations in general traffic levels from one centre to another. In fact, using mean values of $T$ (the driving time scale variable) for the 100 or so examiners in the present sample who had carried out more than 100 tests and adjusting their pass-rates for differences in the average values of $T$, revealed a distribution of ‘corrections’ which had a standard deviation of 2.5 percentage points in pass rate. This ‘correction’ is again virtually independent of the examiner basic pass rate, so that adjusting examiner pass-rates for this route effect did not reduce the variation between examiners. The combined effects of making adjustments for both route features ($F$) and driving time ($T$) on examiner pass-rates resulted in a
distribution of corrections with a standard deviation of 3.9 percentage points – again largely independent of the examiner's basic pass rate.

Weather conditions
Examiners also recorded an assessment of the weather prevailing at the time of the test using a six-point scale: 1 - sunny day/dry roads, 2 - overcast day/dry roads, 3 - not raining but wet roads, 4 - light rain/drizzle, 5 - medium rain, 6 - heavy rain. When this category variable was included in a model which also included age, sex and examiner effects, category 2 (overcast/dry roads) was associated with a pass rate which was significantly above the average ($\chi^2 = 13.2, p < 0.001$), and category 5 (medium rain) was associated with a pass rate which was significantly below average ($\chi^2 = 6.6, p = 0.01$). A re-formulation of these six weather categories into three: 1 - dry weather/dry roads, 2 - not raining but wet roads and 3 - raining, showed that compared to dry roads, pass-rates are a non-significant 2.5 percentage points lower when the roads are wet (but it is not raining) and pass-rates are a significant 3.3 percentage points lower ($\chi^2 = 11.4, p < 0.001$) in those tests carried out in the rain.

The reversing manoeuvres
The study examined the probability of making a serious or dangerous fault associated with each of the three reversing manoeuvres – ie turn in the road, reverse park, and reverse into a limited opening. It found that for the turn in the road, the probability of such a fault was 0.17; substantially lower than the value of 0.29 that applied to the two 'precision' reversing manoeuvres. The implications of this are discussed in Section 5.5.

5.4 Discussion
This study, based on over 20,000 driving tests conducted at a representative sample of test centres in Great Britain, has shown large pass rate variations between some test centres; other variations, not so large but still statistically significant, were associated with driving examiners and some route features and driving conditions. The implications of these variations are discussed in the following paragraphs.

Variation in pass-rates between test centres could arise 'naturally' if, for example, test routes or test conditions vary systematically from one centre to another. Thus in very rural areas where the network density is low and traffic relatively light, pass-rates might be high; in urban centres the reverse might be true. It might also be that test centres in economically deprived areas tend to receive candidates who have not been able to afford to do so much training and practice as those from more wealthy areas.

The route features that seem to be particularly prone to the generation of serious or dangerous faults are pelican crossings and the larger roundabouts; mini-roundabouts did not feature specifically, but were not very numerous either. These features are those requiring drivers to make judgements about giving way, and about taking suitable gaps when presented. It seems reasonable to see the negotiation of these route features as challenging to a new driver and thus providing scope for error. If this finding were to be used to 'standardise' test difficulty, then the difficulty of a route would be most appropriately indexed by the number of pelican crossings and roundabouts on the route.

The proportion of free driving time during the test (as assessed by the examiner) is also a predictor of pass rate – and a better one than the examiner’s assessment of traffic conditions. It is not clear why tests with a high proportion of free driving should have a high pass-rate: in many respects 'free' driving might be expected to be more demanding than the skills involved in driving in slow moving traffic. Perhaps test routes that tend to have less free driving also have features that provoke driver errors – or it may be that the basic control skills involved in stop-start driving are particularly demanding for a significant number of test candidates.

Clearly, differences in test routes within and between centres have a bearing on test standardisation and merit attention. The findings on the importance of free driving time, roundabouts and signal-controlled pedestrian crossings are relevant here. Since this survey was undertaken, DSA has made some changes to test routes to improve coverage of road and traffic conditions. However, full standardisation would be impossible to achieve because of the limited availability of suitable routes. Some within-centre variations between routes would anyway seem acceptable on the grounds that: (a) driving schools tend to prepare pupils for all the routes used by a test centre and (b) candidates are quasi-randomly allocated to test routes, rather than being allocated to easy or difficult routes in a biased and unfair way. The main purpose of the practical driving test is to induce learners to take sufficient training and practice, and the features of the whole set of routes at a test centre will be the main determinant of test standards as perceived by instructors and learners. The fact that some routes may be less demanding than others is likely to have only a second-order effect on the standard that candidates reach before coming for test.

Between-centre differences in routes cannot be justified in the same way, as they would mean that candidates in some parts of the country were being given a less challenging test than candidates in other areas. However, if the routes are typical of conditions met by drivers in the area, the level of safety induced by the test in that area might be as good as the level of safety in areas with more difficult tests. A study of the extent to which drivers restrict their early driving to the area in which they take the test would be instructive.

Further attempts to standardise test routes might involve observing driving only at certain features along the route. This approach has been taken in some Australian and North American states, but brings with it other potential problems. If such a change were to be considered for the British practical driving test, further evaluation would be needed.

The range within which 95% of examiner pass-rates would lie was estimated to be 8 percentage points either side of their test centre’s mean value. For example, in Barnstaple or Aberdeen 95% of examiners would have
pass-rates between 58% and 74%; in Lincoln or Rugby the corresponding range would be from 28% to 44%. The present hierarchical data cannot show the extent to which these variations between centres arise from differences in the competence of candidates coming forward for test, differences in the routes and traffic conditions, or differences between examiners.

The pass rate variations identified in this study clearly merit further investigation, and DSA is conducting studies to develop its understanding of the situation and help improve quality assurance. Results to date indicate, for example, that social deprivation is associated with between-centre variation in pass rate, with centres that have a high proportion of candidates from socially deprived areas tending to have lower pass-rates. It may be that this variable is associated with candidates’ competence via an influence on access to training and practice. The ‘ruralness’ of a test centre (on a scale: rural, small town, large town, outer city, inner city) is also associated with pass rate, with the least rural centres tending to have the lowest pass-rates.

The pass-rates of Contract Examiners were, at the time of the study, significantly higher than those of other grades of examiner. This finding has been overtaken by events, in that DSA no longer has a separate category of Contract Examiners. Nevertheless the finding is discussed below because to do so may be useful in future, to help us understand other between-examiner differences if they occur.

In principle, the higher pass-rates of Contract Examiners could indicate a bias towards Contract Examiners accepting a lower standard of driving. However, other possible explanations also need to be borne in mind. The standard of driving shown on test is depends upon the candidate’s underlying ability, together with factors that help or hinder the demonstration of that ability during the test. An examiner with a high pass rate (once differences between regions, routes and candidates have been corrected for) is not necessarily accepting a worse standard of driving than other examiners. Instead, he or she may be doing something that encourages candidates to perform at their best. This suggests that even though an examiner has a high pass rate, he or she may be applying exactly the same test marking standard as another examiner who has a lower pass rate: there is not necessarily any bias in test marking.

Another possible explanation of the relatively high pass rate of Contract Examiners might be that their candidates tend to be more competent than other candidates. If, for example, Contract Examiners tend to have their schedules arranged after those of other examiners, they will tend to be allocated candidates who are trying to book a test urgently. It seems plausible that such candidates will tend to be people who have recently failed a test and who, by the time they come for retest, might therefore be more competent than the average candidate.

The study identified a number of other factors that need to be taken into account when considering ways that the test could be improved. For example, the examiner selects two out of three standard reversing manoeuvres for each test. The probability of making a serious or dangerous fault associated with the turn in the road was estimated to be 0.17, substantially lower than the estimate of 0.29 for reverse parking and for reversing into a limited opening. It could be argued that this is fair and acceptable, on the grounds that all candidates know in advance the set of manoeuvres that they may be called upon to demonstrate in the test. Nevertheless it does not improve between-test reliability. Since this study was completed, DSA has examined the issue further and has found that choice of manoeuvre is associated with time of day, availability of suitable locations on the chosen test route and other variables. There is clearly a complex interaction going on and it may not simply be the case that one manoeuvre is less demanding than the others.

Drivers taking the test while it was raining had a pass rate about three percentage points lower than those taking the test in the dry. This probably represents an unavoidable difference in test demands, and one that can be accepted on the same basis that differences in route difficulty are accepted. Measures that increase the amount and variety of pre-test driving experience should improve candidates’ capability to drive in the rain. It may also be possible to identify specific items of training (eg practice with switching wipers and demisters on) that would help test performance in the rain.

5.5 Conclusions on factors influencing pass-rates

1 Candidates’ age and sex

As is already well-known, female candidates have lower pass-rates for the practical driving test than males, and older candidates have lower pass-rates than younger ones. These effects need to be taken into account when investigating the effect on pass-rates of other variables (eg time-of-day) that are themselves associated with age and sex of candidates.

2 Day of week and time-of-day

The two earliest test slots on Monday mornings had a statistically significantly higher pass rate than other slots (after adjustment for candidates’ age and sex, and for examiner). This may well be an early morning effect, rather than a Monday morning effect: there were similar effects on other days that did not quite reach statistical significance.

3 Examiner effects

It was estimated that 68% of examiners had a pass rate within four percentage points of their test centre average, and that 95% had a pass rate within eight percentage points of the centre average.

Whatever the candidate’s age and sex, likelihood of passing the test was not found to depend on the sex of the examiner.

Contract Examiners tended to have a slightly higher pass rate than other examiners – a conclusion that is

---

13 There is a further twist to this, in that it could be argued that the ‘better’ candidates are those more likely to be able to maintain their driving performance when stressed. If this were so (and we do not know whether it is) examiners who put candidates at ease would tend to pass ‘worse’ candidates, even though they might be applying exactly the same marking standards as other examiners.
unaltered by adjusting for age and sex of the candidate, though the effect was stronger for female candidates. Further investigation would be needed to discover whether this finding indicates a difference in test standard requiring corrective action. However, the issue has been overtaken by events: there is no longer a separate category of Contract Examiners within DSA, and the quality assurance regime has been further strengthened.

4 Test centre and regional effects
There was a wide variation in pass-rates between test centres – a variation not fully explained by differences in the age and sex of candidates. It is not possible to discover from the survey the extent to which this variation is due to differences in the competence of candidates (eg arising from socio-economic influences on the amount of training and practice they accumulate), differences in the conditions of the test, or differences between examiners. Information on the amounts of training and practice accumulated by candidates at different centres or regions would be useful here. Recent DSA work has shown that test centres with relatively high proportions of candidates from socially deprived areas tend to have lower pass-rates, as do centres in cities and larger towns. These variables may be associated with candidate competence – eg if they affect access to training and supervised practice.

If there are identifiable types of test centre at which candidates come for test particularly poorly prepared, then policies to address the problem may be needed. Low pass-rates at such centres would imply that the test is doing a useful job in screening out such candidates.

Not unexpectedly, there was also variation in pass rate between DSA Areas. Scotland and Wales had higher than average pass-rates; Northern Area, and Midlands and Eastern Area, had lower than average pass-rates. These differences were not explained by differences in candidates’ age and sex. The extent to which they can be explained by candidates’ competence, (itself open to influence by driving instructor competence), or by examiner or test centre effects cannot be established from the present study. A study of the extent to which drivers restrict their early driving to the area where they take their test would be instructive in helping to assess the importance of regional differences in test conditions.

5 Test routes and conditions
Routes with more of the following features tended to have lower pass-rates (sex and age of candidate, and examiner effects having been adjusted for):
- Light-controlled pedestrian crossings.
- Single-lane roundabouts.
- Multi-lane roundabouts.

Routes with 17 such features were estimated to have a pass rate 13 percentage points lower than routes with only one of them.

Routes that included some dual carriageway with three or more lanes in each direction had pass-rates 5.9 percentage points lower than other routes.

The difficulty of traffic conditions experienced on test (as assessed by the examiner in comparison to average conditions at the test centre) was not an important predictor of pass rate.

Tests that involved a high proportion of free driving (as assessed by the examiner) tended to have high pass-rates when candidate’s age and sex, and examiner effects were allowed for.

Tests carried out in the rain tended to have a pass rate about three percentage points lower than those on dry roads. A further examination of the reasons for this might suggest ways of improving candidates’ ability to deal with such conditions.

The effect of the choice of reversing manoeuvres on pass rate is complex, involving interactions between difficulty of the manoeuvre, time of day of test, and possibly other variables. It may not simply be the case that one manoeuvre is less demanding than the others. Nevertheless, the rather large differences between the manoeuvres in terms of the probability of making a serious or dangerous fault mean that this area merits further examination.

6 Test reliability and consistency of candidates’ performance: a test–retest study

6.1 Introduction

DSA examiners are trained with the objective of enabling them to make reliable assessments of candidates’ performance. Examiners are also regularly monitored: a more senior examiner sits in the back of the car and assesses and marks the test independently. The details of the two test assessments are then compared and discussed with the test examiner afterwards. Results from about 350 such check tests were collected as part of this study, and indicate a high level of between-examiner agreement on test outcome. Results from other countries also show that it is possible to achieve high levels of such ‘inter-rater’ reliability in driving tests (eg Keskinen et al., 1988). However, even if examiners’ assessments are perfectly reliable (such that an examiner observing a given test drive would always assess it in the same way, and would agree with the assessments made by all other examiners) there are several reasons to expect that the driving test may have rather limited reliability. For example:

i The test is subject to several quasi-random components such as weather, traffic conditions, specific traffic incidents and choice of test route, each of which may influence the outcome.

ii The test’s marking system makes it vulnerable to some of these components of variation. For a given candidate, the probability that a serious or dangerous fault (ie a failing fault) will be committed during a test will depend on the number of opportunities there are for that fault to occur.

iii As people become more skilled at a task such as driving, their rate of making errors decreases. However, even experienced drivers make errors from time to time. Driving test candidates have accumulated only limited amounts of training and practice, and therefore may still
have a rather high error rate (Groeger and Clegg, 1993; Groeger and Brady, 2004). In other words, if a candidate took exactly the same test several times, facing exactly the same conditions and having exactly the same level of competence each time, he or she would make different numbers of faults on different occasions, passing some tests and failing others.

iv It is possible that candidates will tend to choose just that amount of training and practice that brings them into the range of competence where they have a moderate probability of passing the test. That is, they may decide to come for test when their rate of making errors, together with the other sources of unreliability mentioned above, produce a probability of passing of somewhere around 0.5. This would inevitably lead to low reliability (as measured by test–retest agreements) and low predictive validity14, irrespective of test content and examiner consistency. In short, if most of the people coming for test are on the pass-fail borderline, we cannot expect the test to have high test–retest reliability or predictive validity for these candidates.

Potential sources of unreliability are summarised in Table 7.

<table>
<thead>
<tr>
<th>Component of unreliability</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test conditions (controlled).</td>
<td>Route difficulty, between-examiner effects, day, time.</td>
</tr>
<tr>
<td>Test conditions (uncontrolled).</td>
<td>Weather, roadworks, traffic.</td>
</tr>
<tr>
<td>Test events.</td>
<td>‘Chance’ events during test.</td>
</tr>
<tr>
<td>Within-examiner variation.</td>
<td>An examiner may perform differently on different tests.</td>
</tr>
<tr>
<td>Within-candidate variation.</td>
<td>Even if test conditions were constant, a candidate’s performance as measured by errors actually made during a test may be an unreliable indicator of underlying skill level (eg underlying error rate), so that the candidate would make different numbers of errors on different tests.</td>
</tr>
</tbody>
</table>

To investigate such issues a test–retest reliability study was carried out. This was reported initially by Baugan and Simpson (1999) and is summarised below.

6.2 Study design

Candidates who had already booked but not yet taken a driving test at one of 20 test centres were asked if they would be willing to take a second driving test within a few days of the first. Participants were not told the result of the first test, or given feedback on how they had done, until after they had completed the second. At the end of the second test they were issued with a pass certificate if they had passed either or both the tests. The second examiner was not told the result of the first test until after the second test had finished. Test routes and times were allowed to vary as would happen normally with test bookings, to ensure that these sources of variability were included. For most test pairs it was stipulated that a different examiner would take each test. In five test centres, the examiner allocated was allowed to vary without restriction, yielding 38 test pairs using the same examiner for both tests. Participants did not have to pay for the second test and, if they used a driving school car, the school’s fee for accompanying the candidate to and from the second test was also paid by the study. Following the second test, both the candidate and the instructor were handed a short questionnaire to complete and return. Ideally, from the point of view of the reliability study, candidates would have been discouraged from taking further training between the two tests. This was ethically unacceptable, so the study relied on a short time interval between test and retest to minimise training and practice between tests.

A target sample size of 400 was chosen, since this would enable the proportion of test pairs with agreeing outcomes to be estimated, with 95% confidence, to within +/- 5 percentage points or better. 1048 candidates were invited to express an interest in participating, and 366 actually took part. 279 candidate questionnaires and 319 instructor questionnaires were completed and returned.

At the time the study took place (ie pre-May 1999), candidates did not fail the test for exceeding a pre-set total number of less-serious faults.

6.3 Results

Table 8 shows the main results. 64% 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. Of people who passed the first test, 42% went on to fail the second. Of those who failed the first test, 33% went on to pass the second.

<table>
<thead>
<tr>
<th>Result of 1st test</th>
<th>Result of 2nd test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
<td>Pass</td>
</tr>
<tr>
<td>Fail</td>
<td>154 (42)</td>
</tr>
<tr>
<td>Pass</td>
<td>75 (20)</td>
</tr>
<tr>
<td>Total</td>
<td>229 (63)</td>
</tr>
</tbody>
</table>

Frequencies (% in brackets)

The pass rate for the first test was 37.4% and for the second test was 42.3%. This difference is not statistically significant, but may indicate a learning effect. If so, then without this learning effect the proportion of people passing test 1 who then go on to fail test 2 would probably be higher than the observed 42%. It should be noted that if all candidates were identical in terms of error rates, and came to the first test with a probability of
passing of \( p = 0.374 \) (ie corresponding to the observed pass rate of 37.4\%) and to the second test with a probability of passing of \( p = 0.423 \), this would lead to an expected 51.9\% of agreements between first and second test.

Examiners were asked to rate the level of pass or fail into one of 6 categories: a very good/exceptional pass; a clear, solid pass; a borderline pass; a borderline fail; a definite/clear fail; hopeless, nowhere near the required standard. Table 9 indicates that people with a clear/solid pass or better in the first test were not statistically significantly more likely to pass the second test than were borderline passers of the first test. Similarly, Table 10 shows that people with a definite clear fail or worse in the first test were no more likely to fail the second test than were borderline failers of the first test.

Table 9 Level of pass in first test compared with result in second test

<table>
<thead>
<tr>
<th>Level of pass in 1st test (rated by examiner)</th>
<th>Result of 2nd test</th>
<th>Total (freq.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fail*</td>
<td>Pass*</td>
</tr>
<tr>
<td>Borderline pass</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td>Clear pass or better</td>
<td>60</td>
<td>40</td>
</tr>
</tbody>
</table>

*Cell entry = % of row.

Table 10 Level of fail in first test compared with result in second test

<table>
<thead>
<tr>
<th>Level of fail in 1st test (rated by examiner)</th>
<th>Result of 2nd test</th>
<th>Total (freq.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fail*</td>
<td>Pass*</td>
</tr>
<tr>
<td>Borderline fail</td>
<td>32</td>
<td>68</td>
</tr>
<tr>
<td>Clear fail or worse</td>
<td>38</td>
<td>62</td>
</tr>
</tbody>
</table>

*Cell entry = % of row.

The questionnaire handed to candidates after the second test asked them to indicate how likely they had thought it was that they would pass each test. The rating scale was almost certain to pass; likely to pass; as likely to pass as to fail; likely to fail; almost certain to fail. Overall, 56\% of candidates reported having thought they were likely or certain to pass the first test, 37\% that they were as likely to pass as to fail, and 7\% that they were likely or certain to fail. Table 11 shows how the responses compared with the actual proportions of candidates who passed and failed. It should be noted that at the time they made these ratings, candidates knew the outcomes of both tests. Candidates’ ratings of likelihood of passing the first test were not statistically significantly associated with the actual outcome.

After the second test candidates were also asked to indicate how they thought they had done immediately after the first test. Of those who expressed a view, 52\% reported thinking they had probably or certainly passed, and 48\% that they had probably or certainly failed. Table 12 shows their actual test results.

Table 12 Candidates’ assessment of performance in first test

<table>
<thead>
<tr>
<th>Actual result</th>
<th>Total (freq.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail*</td>
<td>Pass*</td>
</tr>
<tr>
<td>Certain pass</td>
<td>36</td>
</tr>
<tr>
<td>Probable pass</td>
<td>38</td>
</tr>
<tr>
<td>Probable fail</td>
<td>84</td>
</tr>
<tr>
<td>Certain fail</td>
<td>89</td>
</tr>
</tbody>
</table>

*Cell entry = % of row.

Candidates who thought they had probably or certainly passed were statistically significantly more likely to have done so than candidates who thought they had probably or certainly failed (\( p<0.001 \)), although quite a large proportion (around 1/3 to 2/5) of those who thought they had passed were wrong. Candidates who thought that they had probably or certainly failed the first test were generally right. This may have led to a difference in motivation to pass the second test, with people who had failed the first test tending to try harder than people who thought they had passed. Such an effect could have caused the study to overestimate the degree of disagreement between test–retest outcomes, with people who (generally correctly) thought they had failed the first test being strongly motivated to obtain a different result in the second one.

Instructors also were asked to rate their candidate’s likelihood of passing each test. 82\% of candidates were rated as likely or certain to pass the first test, and 17\% were rated as equally likely to pass as to fail. Only 1\% were rated as being likely or certain to fail. Table 13 shows that instructors’ views on the likelihood of a candidate passing a test were associated with the actual outcome, an effect that was statistically significant (\( p<0.001 \)). Pupils whose instructors said they were likely or certain to pass a test were more likely to do so than pupils whose instructors said they had only a 50:50 chance or were likely to fail. There was a general tendency, though, for instructors to over-estimate the likelihood of candidates passing the test. This may suggest that instructors were assessing whether

Table 13 Instructors’ predictions and test outcome

<table>
<thead>
<tr>
<th>Actual result</th>
<th>Total (freq.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail*</td>
<td>Pass*</td>
</tr>
<tr>
<td>Likely or almost certain to pass</td>
<td>57</td>
</tr>
<tr>
<td>As likely to pass as to fail</td>
<td>84</td>
</tr>
</tbody>
</table>

* Cell entry = % of row.
candidates were good enough to pass, rather than whether they actually would pass given all the ‘on-the-day’ factors that might affect performance during the test.

6.4 Explaining test–retest inconsistency

In principle, all the potential sources of inconsistency listed in Table 7 could have contributed to the results shown in Table 8.15. In fact, differences between examiners are unlikely to explain much of the inconsistency – results of the check tests mentioned above showed very high levels of agreement between examiners on test outcome when they observe the same test. 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 it appears that, for a substantial proportion of candidates, 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 a repeat test. This finding is of fundamental importance in helping us to understand how the driving test operates. One view of the driving test considers failers to have demonstrated that their driving is fundamentally different from (ie less safe than) that of passers, and that passers have developed skills and systems of driving that will enable them to deal with ‘chance’ events in a way that test failers have not yet learned to do. Such a model implies a strongly bimodal underlying distribution of individual error rates – such that an individual coming for test has either a low or a high probability of passing.

An alternative model is one in which candidates have a moderate error rate, such that their probability of passing is itself neither very high nor very low. In detecting actual faults, the test in this model is not distinguishing between fundamentally safe and unsafe candidates but maintaining a standard such that candidates with a moderate probability of passing are deemed to be acceptably safe. The test induces people to reach this level of performance before coming for test. This model has a number of implications – eg:

- The pass rate becomes sensitive to minor changes in competence.
- Commission of a single dangerous or serious fault does not necessarily mean that a candidate is a worse driver than one who makes no faults during the test.

On the basis of the results presented here, it appears that this second model is at least partially valid. In fact, the true situation is probably a combination of both models. Some candidates will come for test with a very low, or a very high, probability of passing; others, possibly the majority, will be on the borderline of the required level of competence and have a moderate probability of passing.

15 It could be argued that failing the first test and passing the second does not necessarily imply inconsistency, since candidates’ driving might be expected to improve from test to test: most had a practice lesson just before the second test and the first test would itself have been a learning experience. However, it seems unlikely that small amounts of extra training or practice at this stage of learning would have a substantial effect on error rate and pass rate.

6.5 Improving the reliability of the driving test

As mentioned in Section 2.3, and discussed in Appendix A, despite a pattern of test–retest disagreement like that shown in Table 8, the driving test could still be performing its main function of maintaining standards of training, safety and competence by inducing training and practice. Nevertheless, good reliability is desirable for reasons of efficiency and perceived fairness:

- An unreliable test is likely to be seen as unfair – randomly penalising some candidates and passing others.
- An unreliable test is inefficient and costly, since it would result in unnecessary failures and subsequent retests.
- A test with good reliability and good discriminatory power is likely to be more effective in inducing training and practice.
- An unreliable test would tend to lose credibility amongst candidates and instructors.

Test–retest disagreements stem mainly from the fact that a candidate with a constant underlying error rate will produce different numbers of errors on different occasions. In principle, the problem could be addressed by increasing the numbers of errors observed during the test, which would increase the precision with which the test is able to estimate the underlying error rate. This could be done by increasing the duration of the test or by basing the test on errors that are less severe (assuming that these can be observed reliably) and more frequent.

Appendix B presents the results of a study which gives some useful insights into the effect of individual candidates’ underlying competence on the test–retest reliability of a driving test and on its ability to identify and fail people with unacceptably poor competence. It demonstrates that the likely effects of increasing the duration of the test would be to increase reliability, and to increase the precision with which the test is able to identify candidates with an unacceptably high underlying rate of making errors.

One disadvantage of the above approach is the cost of a longer test. The failure criterion for this type of test can only be an integer number of errors and, if it were to change to two errors, duration would have to be approximately doubled to maintain the current overall pass rate and test standard. A second problem is that the single error which is currently sufficient to produce a test failure is labelled as a serious or dangerous fault: one that involves potential or actual danger. A longer test that failed someone for making two such faults would be more effective than the current test at identifying candidates with a high underlying rate of serious or dangerous faults. However, it would involve passing candidates who make a single serious or dangerous fault. DfT and DSA have made it clear that it would be unacceptable to pass candidates who make such a fault during the test.

An alternative to increasing duration would be to base the test on less severe, more frequent errors – eg the current test’s ‘driving faults’. Potentially, if the test were to fail candidates who make more than a given number of such errors its ability to detect and fail people with unacceptably high underlying error rates could be improved.
Another way of attempting to improve test reliability is to reduce the opportunity for chance events and route/traffic differences to influence the outcome. The probability that a candidate will make a serious or dangerous fault, and thereby fail the current driving test, will depend on the number of opportunities for a fault to occur during the test. This will vary to some extent from test to test (for example, because of varying weather and traffic conditions), introducing an element of unreliability. Possible ways of reducing this include (a) further standardising test routes, (b) assessing performance only on those parts of routes that are standardised and ignoring what happens on other parts of the route, and (c) allowing good performance to cancel-out faults, so that the test score becomes the proportion of times that the candidate gets things right. Tests involving such elements have been introduced in the USA and Australia (e.g. McKnight, 1992). Arguments sometimes raised against introducing such tests in Britain include the difficulty of justifying passing a candidate who has just made a serious or dangerous fault, and a concern that a highly prescriptive and standardised test may reduce the opportunity for examiners to use their skill and judgement in assessing the test drive. There is also the point that drivers do have to be able to deal with unexpected situations. Reducing the opportunity for such situations to influence the test outcome will also reduce the ability of the test to assess this important aspect of driver competence. Note that tests of the type described above will still have poor test–retest agreement if candidates come forward with only a moderate probability of passing. However, if they are successful in reducing chance effects, they may improve candidates’ perceived control over test outcome, and thereby encourage people to reach a higher probability of passing before coming for test.

Tests involving examiners’ overall judgements of skills or competencies have been introduced in, for example, Sweden (Mattsson, 1999), Western Australia (Drummond, 2000) and the Netherlands. In principle, this approach may avoid some of the potential disadvantages of tests based on observing errors, but it may also introduce its own difficulties. These are associated with the reliability with which examiners are able to make the necessary judgements and the fact that it may be less easy for examiners to defend a test result based on such judgements rather than on identifiable observed faults. More evidence is needed here. There is also the point that the examiner’s ratings are themselves presumably based to an extent on the errors observed during the test, so that the two approaches may not in the end be so different from each other.

The above approaches to improving consistency are to do with (a) improving the ability of a test to estimate a candidate’s true expected error rate and thereby distinguish between candidates of moderate competence, or (b) reducing the influence of ‘chance’ circumstances on test outcome. Another approach would be to persuade people to delay taking the test until their driving skills have improved – i.e. until their error rate is lower (driving more consistent). It is therefore important to understand how people decide when to come for test. People might see it as worth attempting the test early, knowing that they will probably fail, but knowing also that if they do pass they will not have to pay for any more driving lessons. If this is going on, the question arises of what would happen if the consequences of failing the test were made more severe. If the cost of a re-test were increased, or if there were a longer compulsory delay between test and re-test, would people then delay coming for test until they were more sure of passing it? If they would, test pass rate would increase, test–retest consistency would improve, the standard of novice drivers would improve, and numbers of retests would decrease.

Section 7 describes the results of a small study designed to gain some insight into decisions on when to take the test.

### 6.6 Conclusions on test reliability

1. Test–retest reliability is low for current candidates for the British practical driving test. Although there are several potential sources of unreliability it appears that inconsistent performance on the part of the candidates themselves is the dominant component.

2. This inconsistent performance can be understood in terms of the candidate’s underlying error rate. A substantial proportion of candidates appear to come for test with an error rate around the test threshold – giving them a medium probability of passing and ensuring low test–retest reliability for current candidates.

3. In itself, 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. Clearly, though, the fact that novice drivers do have an elevated accident risk suggests that more or better training and practice is desirable.

4. A simplified model of the test can give some useful insights into how the test’s reliability, and its ability to identify and fail candidates with unacceptable underlying error rates, may be influenced by test duration, the population distribution of error rates, the severity/frequency of errors included, and the pass-mark and target error rate. Lengthening the test, or basing it on less serious, more numerous errors, should in principle be beneficial – though there are practical difficulties.

5. There would be many benefits if candidates could be persuaded to delay coming for test until they have a high probability of passing it. Pass-rates would increase, test–retest consistency would improve, the standard of novice drivers would improve, and the number of re-tests would reduce. The scope for achieving such a change needs to be investigated further. It may be difficult for some candidates to reach a high probability of passing even with considerable extra preparation for the test.

---

16 In fact most, if not all, jurisdictions that have introduced this type of test continue to fail candidates who make a single serious fault, thereby removing some of the theoretical advantage.
7 How L-Drivers decide when to take the practical driving test – an exploratory study

7.1 Introduction
In Britain, learner drivers are free to choose when to come forward for their driving test. In an attempt to minimise the number of lessons, the total cost of learning to drive, and to achieve as soon as possible the independence that unsupervised driving gives, it is possible that candidates will tend to have just that amount of training and practice that brings them into the range of competence where they have a moderate probability of passing the test. This would guarantee that the test has low test–retest reliability for these candidates, and may happen at a much earlier level of experience and training than is desirable for competence and safety. Potential solutions could include (a) influencing the decision criteria of candidates so that they do not come for test until they have a high probability of passing, or (b) changing the test severity so that candidates with a moderate probability of passing are safer and more competent than they are at present. Solution (a) should also improve both pass rate and test reliability.

Some exploratory interviews were conducted to obtain insight into the factors affecting learner drivers’ decisions about when to take their driving tests, and the scope for influencing these decisions by measures such as increasing the minimum delay between failing one test and taking another – ie by increasing the penalty for failing a driving test.

7.2 Method
Eight driving instructors, and sixteen candidates who had applied for a test but had not yet taken it, were interviewed in September 1998.

Standard ‘unstructured’ exploratory interviewing techniques were used, specially trained interviewers being provided with a topic guide rather than a pre-defined questionnaire. Instructors were asked how their pupils decided to take their test and the extent to which they recommended when their pupils should take their tests. Test candidates were asked why they wanted to drive, and the benefits of being able to drive, about the training and practice they had received, their perceived driving competence, and how they decided when to take their test. Instructors and candidates were asked about certain measures which might affect the decision on when to take the driving test. These included:

- A longer enforced wait between the first test and subsequent re-tests. (Currently candidates can sit a retest after ten days but, in practice, most re-tests are delayed for much longer than this – eg because of test waiting lists or because candidates decide they need longer to prepare themselves for a retest).
- An increase in the test fee.
- An enforced delay between obtaining a provisional licence and taking the test.

7.3 Results
When considering the findings of the interviews it must be remembered first that the sample size was small, as is typical with this type of exploratory study (8 driving instructors and 16 test candidates). Secondly, while efforts were made to recruit participants from a range of backgrounds it is likely that people prepared to give their time for the interviews were the more conscientious instructors and candidates, keen to ensure driving standards are high. It should also be borne in mind that exploratory, qualitative, interviewing does not ask exactly the same questions of each participant. This means that responses should not usually be aggregated and reported as frequencies and percentages. The technique is intended to give insights into the subject area – insights that may need to be further tested by larger scale studies designed to produce quantitative data.

Within the above limitations, the interviews indicated the following:

Most candidates said they decided when to take their test on the advice of their instructor. For some, this decision was made when the candidate felt ‘ready’, or the instructor felt the candidate was ‘ready’. Others booked the test for some time in the future, by which time it was felt that the candidate should be ready; if he or she turned out not to be, the test could be cancelled. A third approach was to choose a test date in the future and simply take whatever lessons or practice needed to reach the required standard by the time of the test. One instructor said that occasionally he would, at their own request, enter candidates for the test when he felt they were unlikely to pass. He did this on the grounds that otherwise they would simply transfer to another instructor who was willing to enter them.

Most candidates (and their instructors) thought that people do not generally take the test before they are ready. However, some appeared to be making a distinction between being ‘ready’ for the test, and having a high probability of passing it – and in fact the perceived probability of passing the test varied greatly from candidate to candidate at the time of booking. Being ready for test seemed to imply being at a stage of competence when things come together, driving becomes ‘natural’, the pupil feels confident (but not too confident) and competent to drive on his or her own, and the instructor can sit back and relax. Being ready for test was seen as giving the pupil a reasonable probability of passing, but the elements of luck in the conditions and events met during the test, test nerves and the perceived foibles of the examiner were believed to intervene in determining the actual test result.

There were indications that some candidates are poor at judging their ability, and thus their readiness for test. Two who had failed their first test said that at the time they had been fairly confident of their ability. However, they now realised after further lessons and practice that they had been less competent than they had judged. Further indications that some pupils were unable to judge their own ability with confidence, come from the fact that when they were told by their instructors that they needed more lessons, they were unsure whether this was true, or
whether the instructors were simply trying to get more business from them. A few candidates were very pessimistic about their ability to pass the test, but said they had been persuaded to take it by their driving instructors. These findings are consistent with those of West et al. (1998), who found that novice drivers’ self-assessments of driving confidence, chances of passing the test, and of how ‘good’ and ‘safe’ they were, did not differ between those who subsequently passed the test and those who failed it. However, West did find that self-assessed driving skill was significantly higher among those who subsequently passed the test.

It appeared that candidates seldom attempted to minimise the total cost of learning to drive by deliberately choosing to come for test before they were really ready. For some, the current cost of the driving test (often taken during a two hour driving lesson which must also be paid for) was a deterrent to taking the test without a reasonably high chance of passing. Those with financial problems tended to say they gave up lessons when they could not afford them and resumed them later, rather than rush to take the test too early on the off-chance of passing.

A suggestion that the cost and delays of tests could be increased as a way of influencing decisions on how much training and practice to have before coming for test was met with hostility by both instructors and candidates. The consensus was that each type of measure was unfair on people struggling to pay for driving lessons or who fail a test for a relatively minor reason or because of nerves on the day. Participants thought that such measures might actually increase stress-related test failures. Enforced delays for the first test and subsequent tests were felt unlikely to increase the amount of tuition or practice before the first test or between tests. Participants thought it more likely that lessons or practice would be taken just before the test or retest rather than throughout the waiting period. Any break in lessons or practice was thought to be detrimental to driving standards overall.

As mentioned above, some candidates and instructors believed that test outcome is determined to a significant extent by factors other than the ‘readiness’ of the candidate – so that a candidate can be ready for test but not have a particularly high probability of passing it. This implies that candidates’ perceived ability to increase their chances of passing by taking more lessons or practice may be rather low. Indeed the interviewers, who had been briefed to explore this directly by asking participants to estimate how many extra lessons would be needed to increase substantially the probability of passing, found that the question was seen as nonsensical. Again this suggests that manipulating the costs and delays of re-testing may not be effective in persuading candidates to take more lessons and practice before coming for test.

A few candidates and instructors did think that increasing costs and delays of testing might deter candidates from putting in for the test too soon. An enforced delay between tests, to allow candidates to reflect and obtain some more practice, was welcomed by some. The preferred delay varied from a couple of days to two or three months, and it was suggested that the length of wait

might depend on the severity of the failure. A delay between obtaining a provisional licence and taking the first test was also thought to be a good idea by some, the recommended period ranging from three months to a year.

Perceived control over test outcome should improve if a way can be found to reduce the perceived importance of test ‘nerves’. One suggestion here was examiner training: several instructors and candidates mentioned that it was easier to pass the test at a particular test centre than at other nearby centres. The examiners were said to be more friendly and to put candidates at their ease; though this was not thought to mean that the standard of pupils passing there was any lower than elsewhere. Reducing the importance of the test itself was also suggested. Some form of continuous assessment of a set syllabus using log books, whereby candidates knew they had successfully completed the set course before the test, might, it was said, serve to boost the confidence of those nervous of the test itself. It might also help to raise tuition standards overall, and prevent candidates entering the test too early in their training. Mock tests with an instructor unknown to the pupil were also suggested as a way of reducing nerves during the actual test, since candidates would have more idea what to expect in the test itself. A two-stage test (separating the manoeuvres from general driving) was also suggested as a way of reducing test stress.

A longer test was generally approved of by both instructors and candidates (although the candidates hoped this suggestion was not implemented before they themselves had passed the test!). Some instructors felt that increasing the time spent during the test on general driving would increase the chances of encountering hazards and allow driving in a greater variety of traffic and road conditions.

It was suggested by instructors and candidates that passing the test did not ensure a driver was good or safe. A candidate may be competent, and drive faultlessly during the test, but then become an aggressive and/or an unsafe driver after passing the test. Instructors and candidates felt this to be a particular problem with young males, keen to show off in front of their friends. The solution (according to instructors) was largely a question of changing attitudes through education and publicity. In particular, a change in advertising by car manufacturers from emphasising speed and power to emphasising safe and environmentally friendly cars was suggested. Other suggestions given by instructors were to limit the size of engine for novice drivers, the number of passengers they may carry (to reduce encouragement to speed in front of peers), and to restrict their night driving for some period of time. It was acknowledged, however, that such measures might be difficult to enforce. Post-test training, such as the Pass Plus course or motorway lessons, were thought to be a good idea by many. It was recognised that people who might benefit most from such courses might be least likely to take them.

7.4 Discussion

The findings of this small exploratory study imply that the reliability and pass-rate of the practical driving test may be
improved by reducing the stress of the test, and reducing the influence of chance events. Such changes should also increase candidates’ perceived (and actual) control over test outcome, and thus make it more worthwhile for them to delay coming for test until they have accumulated more training and practice than they do at present, and have thereby achieved a higher level of competence and a reduced error rate. However, it could be argued that driving itself can be stressful, that a good candidate ought to be able to cope with the stress of a driving test, and that the way to make the test less stressful is for candidates not to take it until they are more competent.

The scope for reducing error rate and increasing probability of passing will be reduced if, by the time candidates come forward for test, their rate of change of error rate with practice is small. In other words, if the rate at which candidates’ performance improves with practice is low, then the scope for improving the probability of passing the test by taking realistically achievable amounts of extra training and practice will also be low. This possibility is given some empirical backing by the work of Groeger et al. (1997). The rate at which many skills are learned is known to be a power function of amount of practice: the rate of improvement with practice flattens-off as the amount of practice increases. Groeger et al., showed that this power law holds for learning to drive: the number of errors that an individual makes on an observed drive is a power function of amount of practice as measured by total miles driven. The power function derived by averaging 35 of Groeger et al.’s participants indicates that the average rate of change of error rate with practice was indeed rather small for people who had accumulated 20 or more hours of practice. These results were not from a representative sample of learner drivers. Moreover, they applied to all errors, including the less severe ones, and may not necessarily apply to the driving test’s serious and dangerous errors. This finding is consistent with the idea that the power function of amount of practice as a function of error rate with practice is low. However, it is at least possible that it may have a stronger effect on the serious and dangerous errors that largely govern the outcome of the driving test. And some ‘higher order’ skills, such as those to do with anticipation and hazard perception, may be at a much earlier stage of development at the time of the test and be improving more rapidly with practice. The safety value of increased pre-test training or practice is demonstrated by some of the results reviewed in Section 10. The fact that drivers learn during the early months of post-test driving is demonstrated by their sharply reducing accident liability during this period.

2 Being ready for test was seen by some as giving a learner a reasonable probability of passing, but elements of luck in the conditions and events met during the test, test ‘nerves’, and the perceived foibles of the examiner were felt to intervene in determining the actual test result.

3 The fact that at least some candidates do not perceive themselves as having much control over test outcome may already act as a disincentive to take more lessons and practice before coming for test. It would be expected to reduce the effectiveness of measures designed to increase training and practice by increasing the penalty for failing a test. Such measures might include increased test fees or increased delays between tests. The indication here is that they would be ineffective because candidates do not think that taking more training or practice will have much effect on their likelihood of passing. Research is needed to explore these issues further, with the aims of:

- testing whether the indicated lack of perceived control can be generalised beyond the small sample used in this study and, if it can,
- attempting to find ways of increasing candidates’ perceived control over test outcome.

4 The scope for increasing the probability of passing by taking more training and practice would be limited if the rate of change of error rate with practice is low. Research is needed to explore further the effect of increased lessons and practice on test performance and error rates, especially for serious and dangerous errors.

5 If it is true that candidate’s perceived control over test outcome is low, attempting to influence decisions about when to come for test through increased fees and delaying retests may be unsuccessful. Improving perceived control over test outcome may change this. The interviewees’ unfavourable views about the fairness and effectiveness of increased fees and delays also need to be taken into account – though such views may themselves be connected with the lack of perceived control over test outcome, and might become more favourable if perceived control can be improved.

7.5 Conclusions on candidates’ decisions about when to take the test

1 Candidates interviewed had not generally come for test before they felt ready.

Note that Groeger et al.’s results should not be taken to imply that drivers stop learning after about 20 hours of lessons or practice. They do indicate that for the group studied, a small amount of extra training and practice would not have much effect on the ‘all-error’ rate. However, it is at least possible that it may have a stronger effect on the serious and dangerous errors that largely govern the outcome of the driving test. And some ‘higher order’ skills, such as those to do with anticipation and hazard perception, may be at a much earlier stage of development at the time of the test and be improving more rapidly with practice. The safety value of increased pre-test training or practice is demonstrated by some of the results reviewed in Section 10. The fact that drivers learn during the early months of post-test driving is demonstrated by their sharply reducing accident liability during this period.

8 Novice driver accidents

8.1 Introduction

One of the main justifications for seeking improvements to the driving test is the elevated accident liability of novice drivers. If it could be shown that novice drivers were particularly susceptible to certain types of accident, this might suggest deficiencies that could be addressed by improved training and testing. Even if an excess of a particular type of accident could be attributed to novice drivers’ exposure to risk, rather than to a skills or knowledge deficiency, it might in principle be possible to use training and testing to reduce the accidents. The project therefore included a review of what is known about the characteristics of novice driver accidents. The review, published as Maycock (2002), covered the information
8.2 Summary of findings

- International accident statistics show that accident rates of younger drivers are, in relation to the relevant population, over twice those for older drivers. And the youngest drivers have the highest risk of accident involvement per mile driven.
- When statistical modelling is used to adjust for the effects of mileage and other variables, the accident liability of novice drivers is seen to fall sharply during the first few years of driving. Both age (maturity) and experience play important parts in this decrease.
- Young novice drivers appear to be more at risk than older, more experienced drivers in all or almost all situations and types of accident. Even when, relative to older drivers, young novices have a low proportion of a given type of accident, their absolute risk for this type of accident will usually be much higher than for the older, more experienced drivers.
- Compared with older more experienced drivers, novice drivers have a greater proportion of their accidents at night, on bends, at inappropriate speed and with no other vehicle involved. The excess of novice driver accidents of these types may be as much to do with exposure, type of journey, and motivational factors as it is to skills deficits, but this does not decrease their importance.
- The difference in the accident rates of male and female drivers depends on the way these rates are estimated. In terms of accident rates per mile, the accident rates of women drivers can be as high or even higher than that of men. However, when statistical modelling is used to adjust for the difference in the annual mileages driven by male and female drivers, the accident liabilities of novice women drivers averaged over the first three years of driving are about 12.5% lower than those of novice male drivers – a difference that reduces as age increases. Older women drivers have much the same accident liabilities as male drivers with the same exposure to risk.
- For both male and female drivers, the proportion of accidents that take place on motorways and dual carriageways, and in built-up areas, increases with age; the proportions on single carriageways and in non built up areas decrease with age. Exposure data here are not robust but it appears that, especially for male drivers, these effects are not fully explained by changes in the pattern of exposure with age.
- Compared with older drivers, accidents involving pedestrians form a relatively low proportion of young drivers’ accidents. This does not appear to be explained by young drivers doing relatively low proportions their driving in urban areas. On the contrary, what evidence there is indicates that the proportion of time spent driving in urban areas actually decreases with age. Therefore, it appears that the novice driver accident excess is not disproportionately associated with difficulties in coping with pedestrians.
- Young drivers, both male and female, aged 17-19 have about twice the proportion of their accidents on bends as drivers aged 30-39. They also have a greater proportion of their accidents while turning right at crossroads.
- Though they have a slightly lower proportion of accidents at roundabouts than older male drivers, the roundabout accidents that younger male drivers do have are more likely than those of the older drivers to involve ‘going ahead’. Those of the older group are more likely to involve ‘waiting to go ahead’.
- For young female drivers, a relatively high proportion of their roundabout accidents involve turning right, and a relatively low proportion involve waiting to go ahead.
- The proportion of accidents having speed as a contributory factor declines with age.
- Young drivers in a Californian study had a much higher proportion of accidents involving poor visual search than did older drivers.
- Young drivers, particularly young male drivers, have a higher proportion of ‘active shunts’ (ie running into others) and loss of control accidents than older drivers.
- The proportion of injury accidents involving alcohol is no higher for the youngest drivers than it is for other drivers up to age 40. The proportion is somewhat higher for the 20-24 year old group, however.
- Young drivers are more likely than drivers in their middle years to be judged by themselves and investigators as being responsible for their accidents.
- Young novice drivers regard breaking the speed limit and having the wrong attitudes as not being particularly important as causes of accidents; older novices see them as more important. Novices of all ages see driving too fast for the conditions as an important cause of accidents. These findings imply that the youngest novice drivers do not regard the speed limit as a useful indicator of safe speed.
- Young novice drivers are more worried about driving in fog, snow and ice than older novice drivers.
- Young novice drivers are more likely than older novices to say they disregard speed limits, exceed speed limits in build up areas and indulge in competitive behaviours such as trying to beat other drivers away from traffic lights or getting involved in unofficial races. Drivers who indulge in such ‘violation’ behaviour tend to have a raised accident liability.

8.3 Implications for the practical driving test

The main relevance of the overall effect of experience on novice driver safety is not so much to indicate how the driving test could be improved but, rather, to suggest that there is scope for improvement. As discussed in previous sections the effect shows that novice drivers learn some useful things during their early unsupervised driving. If
that learning could take place in relative safety during pre-test driving\textsuperscript{18}, there could be a substantial road safety benefit. Thus, the experience effect suggests there is potential for improving the way that people learn to drive; the driving test itself, as has been argued, one way of influencing this.

The effect of age (maturity) alone on accident liability might, at first glance, seem to provide a basis for assessing the driving test. If the driving test is an effective means of identifying people who would go on to have a raised accident liability, perhaps it ought to be capable of detecting that older novice drivers are safer than young novice drivers. In fact, the pass rate for older drivers is significantly lower than that of the younger drivers (eg Forsyth, 1992 and Section 5 of this report), which could be taken to suggest that the test is not performing well in this regard. However, as pointed out in Section 2, a low pass rate for a group of candidates does not necessarily imply anything about the driving competence of those in the group who actually do pass. It may be that older people, in tending to take more tests than younger ones, build up useful experience and practice such that by the time they do pass they are better, safer drivers than the younger candidates who tend to pass with smaller amounts of training and experience and smaller numbers of retests. It is also possible that the extra experience and practice built up by older learner drivers tends to produce a more stable driving style, less susceptible to degradation during the early months of unsupervised driving.

A related point, discussed in Section 2, is that a driver’s competence as revealed during a driving performance test may not be a good indicator of his or her subsequent unsupervised driving behaviour. The relative safety of older novice drivers may have at least as much to do with the way they choose to behave on the road as it does to their driving ability. This interpretation would be consistent with the indications that young people – especially young males – tend to be faster and more aggressive drivers, more likely to get involved in ‘active’ accidents, and with a greater propensity to violate formal and informal traffic rules, than their older counterparts. These characteristics are difficult to detect (or perhaps easy to conceal) on test. The question of whether it might be possible for the driving test to include items designed to assess the attitudes and motivations that predict unsupervised driving behaviour is discussed in Section 3.3.9.

The findings concerning accidents at night, on bends and overtaking, seem likely to be particularly susceptible to the difference between supervised test performance and unsupervised driving behaviour. This is especially so with regard to what are perhaps the most important aspects of speed choice – ie judging the speeds appropriate on higher speed roads at times when speed is not constrained by the presence of other traffic. However, test routes could be made to include more bends on faster roads and the May 1999 modifications to the test included a move in this direction. In principle, attitude and knowledge questions built into the practical or theory tests might help to emphasise the importance of such behavioural issues, encouraging improved coverage of them in driver training and reducing the degree to which drivers’ behaviour departs from good, safe behaviour once they are unsupervised. However, as argued in Section 10 it may be asking too much of a driving test alone to meet the requirements of an effective driver training/testing regime. It may be more appropriate to address the finding that young drivers are over-represented in night-time single-vehicle accidents on bends by means of other measures designed to influence driver training, or to modify driver behaviour.

From time to time concerns have been expressed about the omission of motorway driving from driver testing and pre-test training. However, the findings of the review indicate that, although young novice drivers are more at risk than other drivers in all or most situations, motorways do not present them with a disproportionately higher risk than other road types. If anything, the figures suggest that it is national speed limit single carriageway rural roads that give particular problems to young novice drivers. In principle, it would be possible for test routes to be modified so they give better coverage of such roads – which should induce more training and practice on them. Indeed, improvements in this direction were made in the May 1999 changes to the test. In practice, the extent to which this can be achieved at reasonable cost is limited. Again, it might be necessary to use means other than the driving test to encourage learner drivers to accumulate more training and practice on rural roads.

The findings suggest that it could be beneficial to give learner drivers more training and experience in judging appropriate speeds for bends and roundabouts, as well as in negotiating junctions and overtaking. Again, though, the scope for achieving this by adjusting the content of the driving test is limited by availability of routes with suitable features and where speeds are not determined by other traffic. Also, speed on bends, and deciding when to overtake, seem likely to be particularly susceptible to the difference between supervised test performance and unsupervised driving behaviour. Here, too, the driving test probably needs to be supplemented by other measures to improve training and experience, and control risk during early solo driving.

The indication that young or novice drivers have inferior ‘search’ strategies is addressed partially by the introduction of a computer-based ‘hazard perception’ test. Such perceptual abilities are rather difficult to measure in a practical test.

\textsuperscript{18} Or during post-test driving with risk being controlled by probationary restrictions – see Section 10.
Self reported formal and informal traffic violations have been shown to be strong predictors of accident liability in a number of studies. Although the precise nature of this link remains unclear, it seems highly desirable to reduce drivers’ propensity to depart from good, safe, driving behaviours. The practical driving test itself does not seem to be a particularly appropriate tool here, though using the theory test to tap knowledge of the importance of such behaviours may be useful. Assessment of attitudes during theory or practical tests, and feedback on their importance, may also be useful. However, directly influencing driver training to emphasise the risks of such behaviour and change attitudes towards it may be more appropriate than relying only on modifying the driving test itself – as may measures to penalise violational behaviour during early post-test driving.

As regards their implications for the driving test, gender differences in accident liability parallel the age differences. First, the fact that there are gender differences suggests that there would be safety improvements if young men drove more like young women do. In principle, the training/testing system could have a part to play in achieving this, though the difference in safety between young men and young women drivers is probably more to do with attitudinal/motivational effects rather than differences in skill. Secondly, it might at first sight seem that a ‘good’ test ought to be able to distinguish the relatively safe driving of women from the relatively less safe driving of men. In fact female candidates have a pass rate about 5-10 percentage points lower than that of male candidates at all ages (Section 5.3.1), which might be interpreted as counting against the test in its present form. Such an interpretation would, however, fail to recognise the complexity of the situation. As argued in Section 2.3, the fact that a group of candidates has a low pass rate tells us little about the probability of passing or the competence of those in that group who do pass.

9 Faults made during the test, and accidents in the first six months of driving

9.1 Introduction

As described in Section 3, the British practical driving test involves candidates being continuously monitored for 48 categories of predefined errors, each of which is assessed as a dangerous, serious or less serious fault. Serious or dangerous faults are those judged to involve potential or actual danger respectively, and candidates making one or more of them fail the test. Before May 1999 there was no specified limit on the number of less serious faults a candidate was allowed to make without failing. Only the first two less serious faults in any category were recorded. Since May 1999 examiners have recorded all less serious faults, renamed ‘driving faults’ and candidates making 16 or more of them have failed the test. Setting the maximum permitted number of driving faults at 15 has meant that just under 0.5% of candidates failed the new test through making too many driving faults. The intention was to tighten the failure criterion if it could be shown that to do so would fail candidates with raised accident risk.

To provide evidence on whether tightening the failure criterion would be justified, the project investigated the relation between the number of driving faults recorded during the test, and accidents in the first six months of post-test driving.

It was recognised from the outset that the ability of driving faults to identify drivers who go on to have a raised accident liability would be limited for the reasons described in Appendix A. Nevertheless, if an adjustment to the test were able to identify and fail a new group of drivers with an only slightly raised accident liability, the potential accident savings might be worthwhile. Initially the failers would be returned for more training or practice before being allowed to drive unsupervised; in the longer term the new testing requirements might be expected to induce learners to reach a higher standard before coming for test in the first place. The indications from the first TRL Cohort Study (Maycock and Forsyth, 1997) were that the less serious faults might have useful predictive power.

The question of whether it is possible to identify a group of relatively unsafe drivers by means of driving test faults is conceptually not a simple one. A group of drivers who have more accidents in a given time than another group might well be regarded as relatively unsafe, and as a worthy target for a safety intervention. But accident liability depends on many variables and, crucially, on exposure to risk. It is therefore possible that the raised accident liability of the group might be explained by such variables. This would imply that although the group members had more accidents than other drivers, their standard of driving was not intrinsically worse or less safe. Corrective measures (such as inducing further training and practice by causing the high accident group to fail the driving test) might produce a safety benefit for the group, but the fairness of applying the measures solely to group members when they are intrinsically just as good as other drivers might be questioned.

Consider now a group whose members’ driving is intrinsically worse/less safe than that of the rest of the driver population: they have a raised accident liability when exposure to risk is controlled for. This group, too, might be regarded as a good target for safety interventions such as using the driving test to encourage them to have more training and experience. However, in practice the group members might actually have no more accidents than the rest of the population, because their exposure to risk is lower. It might then be argued that targeting a safety intervention on the group was unfair to the group members, or inefficient in terms of resource allocation, because their relatively unsafe driving was already compensated for by the reduced exposure to risk.

The implications of the above situations for safety interventions are not clear-cut. In each situation, interventions might be effective in reducing the numbers of accidents amongst the high-accident group, or the unsafe-driving group, but might be seen as unfair – in the one case because they penalised a group whose standard of driving was not inferior, and in the other because they
penalised a group whose inferior standard of driving was compensated for by a reduced exposure. Only where a group of people drive less safety than the rest of the population and actually have more accidents, does the situation become more straightforward to deal with. Clearly, therefore, it was necessary for the survey to examine the differences in actual numbers of accidents between high and low faults groups, and to investigate whether there is a relation between faults and accidents once any differences in exposure between high and low faults groups have been allowed for.

9.2 The survey
9.2.1 Sample design
Analysing a sample of driving test records showed that only 1.5% of test passers made 15 driving faults so it was decided to pool those making 14 and 15 faults into one group. Similarly, people making 0, 1 and 2 faults were put into one group, leaving thirteen 'number of driving faults' groups in total. The smallest of these was the 13 faults group, accounting for about 4.4% of test passers.

Names and addresses of drivers who had passed the practical driving test were obtained from DSA together with their driving test records. The survey sample was issued in monthly waves, to people who had passed the test about seven months earlier. The first wave had equal numbers of people in all 13 groups, as had waves two and three. Two further waves included only people making eight or more driving faults. The size of each wave was limited by the number of candidates passing the test each month who made 13 faults (all these candidates were included in the sample).

9.2.2 The questionnaire
The questionnaire asked about accidents in the first six months of post-testing, pre-test driving experience, driving exposure since passing the test, near accidents, and police warnings and prosecutions. It included a number of questions about driving ability, driving style, attitudes to driving violations, hazard involvement, and aberrant driving behaviours that previous research has found to be related to accident liability.

9.2.3 Fieldwork
The survey was administered by the National Foundation for Educational Research (NFER) under contract to TRL. A single reminder, including a copy of the questionnaire, was issued to people who had not responded two weeks after the initial posting. The questionnaire title was ‘Transport Research Laboratory Driving Questionnaire’ and the survey was described as a survey of the driving experiences of recently-qualified drivers, being conducted to help improve driver testing.

9.3 Results
9.3.1 Achieved sample
The response rate to the survey was 42.2%. Some respondents were excluded – eg because they provided incomplete data for variables that were to be used in statistical modelling. This left 29,559 records for the main analyses, and reduced the effective response rate to 37%.

9.3.2 Proportions of drivers involved in accidents
Figures 7 and 8 show 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. There was, however, a suggestion of such a relation for women. This was confirmed by a weighted regression analysis of the proportion of each driver fault group who were involved in accidents. This showed, for women, a weak linear component that was statistically significant at the 10% level though not at the more usually specified 5% level.

The above analyses were in terms of the proportion of people in each number-of-faults group who were accident involved. Parallel analyses were carried out on the mean number of accidents reported by the members of each group. They lead to similar conclusions.

9.3.3 Can a driving faults threshold identify a group who go on to have more accidents than other drivers?
The following analysis first sets the simulated ‘failure’ threshold at 14 driving faults and compares the accident involvement of two groups of people: those with 14 or 15 faults and those with less than 14. Next, the threshold is set at 13 faults, and the accident involvement of those making 13 or more faults is compared with those making fewer than 13. This is repeated for failure thresholds down to 7 faults. Tables 14 and 15 show the results.

The tables show that women who make 11 or more driving faults were statistically significantly more likely to report an accident than the rest. The same applies for 10 or more, 9 or more and 7 or more faults – in other words, setting the failure threshold at 11, 10, 9, or 7 would fail women with a higher likelihood than the rest of reporting an accident. Setting the failure threshold at 14, 13, 12 or 8 driving faults would tend to have the same effect though the effects for these groups are not statistically significant – note, though, that the statistical power of the comparisons decreases in the upper rows of the table because the sample size for the higher faults group diminishes.

These differences in accident involvement are rather small: for example, 13.4% of women who made 11 or more driving faults reported an accident and 12.2% of the rest did – a difference of 1.2 percentage points. The 95% confidence interval for this difference is 1.2 ± 1.07 percentage points. This is equivalent to the statement that women making 11 or more driving faults were (9.8 ± 8.8)% more likely than the rest to be involved in an accident

For men, it is not possible to find a pass-fail threshold that fails a group of ‘higher accident’ men. There is even a (non-significant) tendency for men who make 14 or more
Table 14 Accident involvement of high and low faults groups – males

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Higher faults group</th>
<th>Lower faults group</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accident involved</td>
<td>Not involved</td>
<td>Proportion</td>
</tr>
<tr>
<td>14,15 v rest</td>
<td>179 939 0.160</td>
<td>2156 10340 0.173</td>
<td>ns</td>
</tr>
<tr>
<td>13-15 v rest</td>
<td>378 1898 0.166</td>
<td>1957 9381 0.173</td>
<td>ns</td>
</tr>
<tr>
<td>12-15 v rest</td>
<td>578 2852 0.169</td>
<td>1757 8427 0.173</td>
<td>ns</td>
</tr>
<tr>
<td>11-15 v rest</td>
<td>801 3877 0.171</td>
<td>1534 7402 0.172</td>
<td>ns</td>
</tr>
<tr>
<td>10-15 v rest</td>
<td>1014 4884 0.172</td>
<td>1321 6395 0.171</td>
<td>ns</td>
</tr>
<tr>
<td>9-15 v rest</td>
<td>1241 5915 0.173</td>
<td>1094 5364 0.169</td>
<td>ns</td>
</tr>
<tr>
<td>8-15 v rest</td>
<td>1492 6986 0.176</td>
<td>843 4293 0.164</td>
<td>ns</td>
</tr>
<tr>
<td>7-15 v rest</td>
<td>1609 7631 0.174</td>
<td>726 3648 0.166</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 15 Accident involvement of high and low faults groups – females

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Higher faults group</th>
<th>Lower faults group</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accident involved</td>
<td>Not involved</td>
<td>Proportion</td>
</tr>
<tr>
<td>14,15 v rest</td>
<td>188 1205 0.135</td>
<td>1833 12719 0.126</td>
<td>ns</td>
</tr>
<tr>
<td>13-15 v rest</td>
<td>388 2485 0.135</td>
<td>1633 11439 0.125</td>
<td>ns</td>
</tr>
<tr>
<td>12-15 v rest</td>
<td>586 3789 0.134</td>
<td>1435 10135 0.124</td>
<td>ns</td>
</tr>
<tr>
<td>11-15 v rest</td>
<td>793 5124 0.134</td>
<td>1228 8800 0.122</td>
<td>p&lt;0.05</td>
</tr>
<tr>
<td>10-15 v rest</td>
<td>998 6431 0.134</td>
<td>1023 7493 0.120</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>9-15 v rest</td>
<td>1185 7811 0.132</td>
<td>836 6113 0.120</td>
<td>p&lt;0.05</td>
</tr>
<tr>
<td>8-15 v rest</td>
<td>1360 9134 0.130</td>
<td>661 4790 0.121</td>
<td>ns</td>
</tr>
<tr>
<td>7-15 v rest</td>
<td>1488 9949 0.130</td>
<td>533 3975 0.118</td>
<td>p&lt;0.05</td>
</tr>
</tbody>
</table>
faults to be less likely than the rest to have an accident. The same applies to men who make 13 or more, and 12 or more, faults.

Tables 14 and 15 also demonstrate that women are less likely than men to be involved in accidents. This applies in all the ‘numbers of faults’ categories. Women who make (say) 11 or more faults are more likely to report accidents than other women, but they are less likely to have accidents than the men in any ‘number of faults’ category.

### 9.3.4 Age and exposure of high-faults drivers
Table 16 shows that people who made more driving faults tended to be a little older than the rest, to drive fewer miles in the first six months post-test and to be more likely to ‘rarely or never’ drive in the dark. This was true for both men and women.

**Table 16 Age, exposure and number of driving faults**

<table>
<thead>
<tr>
<th>No. of driving faults</th>
<th>Average age (yrs)</th>
<th>Proportion who drive often or sometimes in the dark</th>
<th>Average miles (6 mths)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>0,1,2</td>
<td>20.5</td>
<td>20.7</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>20.0</td>
<td>20.9</td>
<td>0.98</td>
</tr>
<tr>
<td>4</td>
<td>19.8</td>
<td>21.1</td>
<td>0.97</td>
</tr>
<tr>
<td>5</td>
<td>20.3</td>
<td>21.3</td>
<td>0.96</td>
</tr>
<tr>
<td>6</td>
<td>20.6</td>
<td>21.2</td>
<td>0.96</td>
</tr>
<tr>
<td>7</td>
<td>20.5</td>
<td>21.2</td>
<td>0.97</td>
</tr>
<tr>
<td>8</td>
<td>20.5</td>
<td>22.3</td>
<td>0.95</td>
</tr>
<tr>
<td>9</td>
<td>20.8</td>
<td>22.1</td>
<td>0.96</td>
</tr>
<tr>
<td>10</td>
<td>20.9</td>
<td>22.7</td>
<td>0.96</td>
</tr>
<tr>
<td>11</td>
<td>21.1</td>
<td>22.8</td>
<td>0.96</td>
</tr>
<tr>
<td>12</td>
<td>21.2</td>
<td>23.0</td>
<td>0.96</td>
</tr>
<tr>
<td>13</td>
<td>21.6</td>
<td>23.5</td>
<td>0.96</td>
</tr>
<tr>
<td>14,15</td>
<td>21.2</td>
<td>22.9</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Mileage and driving in the dark are indices of exposure to risk, so their associations with numbers of driving faults might be expected to reduce the accident liability of the high-faults groups and help to explain the lack of a strong relationship between numbers of faults and accidents. In other words, high faults people might be intrinsically less safe as drivers, but might have no more accidents than low faults people because they have less exposure to risk. There is a complication here: in this sample, where everyone has completed six months of post-test driving, mileage and proportion of night driving become important indices of experience as well as of exposure. Given that accident liability is known to fall sharply as drivers gain experience, the experience effect is likely to offset at least some of the exposure effect of mileage. The high faults people tend to have lower exposure to risk (reducing their accident liability) and less driving experience (which may increase their accident liability).

### 9.3.5 Fitting accident models
To find out whether a relation between driving faults and accidents emerges once the effects of the above variables are removed, models of the type developed by Maycock et al. (1991) were fitted to the data. First, the number of accidents was modelled as a function of mileage covered in the first six months, age when passing the test, and whether the driver drove often or sometimes at night rather than rarely or never. A term for driving faults was then added to the model. The model has the following functional form:

\[
\text{Accidents} = \text{miles}^{b1} \times \exp\{b2/\text{age} + b3*(\text{drive in dark}) + b4*(\text{driving faults})\}
\]

Males and females were treated separately. Table 17 shows the values of the parameters that were statistically significant at the 5% level or better.

**Table 17 Accident model parameters (all-accident model)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model without driving faults</th>
<th>Model with driving faults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>b1, miles</td>
<td>0.279</td>
<td>0.283</td>
</tr>
<tr>
<td>b2, 1/age</td>
<td>40.48</td>
<td>19.36</td>
</tr>
<tr>
<td>b3, drive in dark = no</td>
<td>-6.502</td>
<td>-4.651</td>
</tr>
<tr>
<td>b4, all driving faults</td>
<td>-5.955</td>
<td>-5.287</td>
</tr>
<tr>
<td>Deviance(^2)</td>
<td>9447.7</td>
<td>9514.5</td>
</tr>
<tr>
<td>Diff. in deviance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of non-Poisson variation explained(^3)</td>
<td>9.6%</td>
<td>13.4%</td>
</tr>
</tbody>
</table>

1 e.g. the model term \(b3*(\text{drive in dark})\) takes the value \(-6.195\) for men who drove often or sometimes in the dark, and \(-6.752\) for those who did not (with-driving-faults model).

2 The deviance statistic indicates how well the model fits the data. Difference in deviance between the models with and without driving faults indicates the improvement in fit produced by including driving faults. It is distributed as chi-square on the number of degrees of freedom change between the models (in this case, one extra parameter, so one d.f. change).

3 For an explanation of this see Appendix B of Forsyth et al. (1995).

4 Statistically significant (p<0.001).

Mileage and proportion of night driving are each positively associated with accidents in the model, indicating that in this cross-sectional study, the exposure effects of these variables outweighed their experience effects.

The fact that the driving faults term is statistically significant indicates that there is a relation between driving faults and accidents not apparent from the data presented in Figures 7 and 8, where it is masked by differences between high and low faults drivers in terms of mileage, driving in the dark, and age. Figures 9 and 10 illustrate the relationship between driving faults and accidents as predicted by the model once mileage, age and driving in the dark are held constant.

### 9.3.6 Practical significance of the modelling results
The models can be used to compare the predicted average accident liabilities of two groups of people – those making \(n\) or more driving faults, and those making fewer than \(n\) faults – while removing the effects of between-group differences in age, mileage and driving in the dark. The results show, for example, that men making 12 or more driving faults would have a mean accident liability
between 7.4% and 25.6% higher than the others. The equivalent figures for women are 9.8% and 31.6%. (These ranges are 95% confidence intervals – i.e. there is a 95% likelihood that the true values lie within the ranges.)

9.3.7 Responsibility for accidents
The foregoing analyses are concerned with all the accidents that drivers reported in the survey. In principle, it is possible that different, perhaps stronger, relations between driving faults and accidents might exist for subsets of accidents. One obvious subset consists of accidents for which the survey respondents themselves were responsible. Respondents were asked how much they were to blame for the accident. There was no relation of any practical importance between number of faults and self-assessed blameworthiness.

9.3.8 Low speed manoeuvring accidents
In a study of this type, by far the majority of accidents involve only minor damage and no injuries. The argument for basing safety decisions on this type of information has always been that a strong element of chance influences the outcome of an accident, so that minor accidents have a similar underlying causal structure to those with serious consequences and therefore have the potential for serious consequences themselves. To the extent that this is true, interventions that reduce the number of minor accidents would also be effective in reducing serious accidents.

While this argument is undoubtedly a strong one, it is obvious that there are some differences in causal structure between minor and severe accidents. For example, accidents at very low speeds are unlikely to result in serious injury to the vehicle occupants. This is not to say that such accidents are unimportant. They gain importance from being numerous, and the behavioural aspects of their causal structure may in fact have much in common with higher speed accidents. The same poor vehicle control or poor observational skills that lead to a low speed accident might also increase the risk of higher speed accidents with severe consequences. Moreover, a low speed accident that involved reversing into an unseen bollard might equally well have involved reversing into a child.

It is therefore not possible to pre-define a set of accidents that are truly trivial, and that tell us nothing about the propensity of an individual to have more severe accidents. Nevertheless, concerns were expressed during the planning of this survey that it would be dominated by trivial accidents having little to do with road safety, and that it might lead to failure criteria being set for the driving test that would unfairly penalise candidates with a raised liability for trivial accidents alone. Low speed manoeuvring accidents were suggested as an example of such accidents. It was therefore decided to ask respondents whether they would describe each accident as a ‘low speed manoeuvring accident’, so that the data could be analysed both including and excluding such accidents.

Excluding such accidents in fact produced results similar to those for all accidents.

9.3.9 Accidents and types of driving fault
All the foregoing analyses concern the total number of driving faults recorded for a test candidate, and are relevant to the question of whether the test failure criterion for total driving faults should be adjusted. In principle, some types of fault may be more important than others as indicators of driving safety. Their association with accidents might have been masked by the presence in the analyses of other, less useful, faults. It was therefore necessary to investigate whether subsets of driving faults are predictive of accidents, with a view to setting separate test-failure criteria for subgroups of driving faults.

In the original TRL cohort study of novice drivers, Maycock and Forsyth (1997) grouped the driving test faults into two main categories, each with four subcategories, as shown in Table 18.
In the present study, statistical modelling was used to find whether each of these eight categories of faults was predictive of accidents once the effects of age, mileage and driving in the dark have been taken into account. The two faults associated with the emergency stop were excluded from the statistical modelling because, since May 1999, only about a third of candidates have been tested for the emergency stop.

Models were produced for all accidents, and for non-low-speed-manoeuvring accidents. To allow the relative importance of each category of fault to be compared, Table 19 summarises the ‘deviance explained’ by each category of faults. The values of deviance explained provide a way of ranking the importance of the fault categories as predictors of accidents both within and between models, once differences in mileage, age, and driving in the dark have been allowed for. Table 19 indicates that:

- Total faults predicted all-accidents somewhat better than it did non-low-speed-manoeuvring accidents. This was especially so for male drivers.
- Total faults were better than any other faults category at predicting all-accidents.
- For non-low-speed-manoeuvring accidents, the best predictors were normal driving faults for male drivers, and total control faults for female drivers.

### Table 19 Fault category parameters – deviance explained by driver fault categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Males</th>
<th>Females</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>All accidents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control faults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c – controls</td>
<td>6.1</td>
<td>6.1</td>
<td>2.1</td>
<td>4.8</td>
</tr>
<tr>
<td>d – normal driving</td>
<td>9.8</td>
<td>7.6</td>
<td>11.3</td>
<td>4.1</td>
</tr>
<tr>
<td>m – manoeuvres</td>
<td>5.8</td>
<td>5.6</td>
<td>5.6</td>
<td>7.3</td>
</tr>
<tr>
<td>s – mirrors/signals</td>
<td>0.0</td>
<td>0.9</td>
<td>0.9</td>
<td>3.7</td>
</tr>
<tr>
<td>Total control faults</td>
<td>10.3</td>
<td>12.9</td>
<td>4.5</td>
<td>17.1</td>
</tr>
<tr>
<td>Perceptual / judgemental faults</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a – awareness</td>
<td>3.9</td>
<td>5.2</td>
<td>1.1</td>
<td>0.7</td>
</tr>
<tr>
<td>v – violations</td>
<td>0.1</td>
<td>0.8</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>f – too fast</td>
<td>6.8</td>
<td>6.4</td>
<td>6.4</td>
<td>0.5</td>
</tr>
<tr>
<td>h – too slow</td>
<td>1.0</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Total perceptual / judgmental faults</td>
<td>8.4</td>
<td>6.6</td>
<td>5.4</td>
<td>1.6</td>
</tr>
<tr>
<td>All faults</td>
<td>17.3</td>
<td>20.6</td>
<td>8.3</td>
<td>16.5</td>
</tr>
</tbody>
</table>

- For male drivers’ non-low-speed-manoeuvring accidents, normal driving faults were a better predictor than total control faults, suggesting that the other components of total control faults – ie controls, manoeuvres, and mirrors/signals – somehow reduce the usefulness of total control faults as a predictor of these accidents.

- Male drivers’ non-low-speed-manoeuvring accidents were less predictable from driving test faults than were the three other categories of accidents. This is consistent with the hypothesis that for men, once low-speed-manoeuvring accidents are discounted, factors associated with behavioural choice rather than performance error are even more dominant as accident predictors than they are for women.

- Mirrors/signals faults, and driving too slowly on test, contributed little or nothing to explaining accident liability.

- Violation faults during the driving test contributed little or nothing to predicting accident liability.

The implications of these findings for the driving test would appear to be that there is little justification for basing failure criteria on the subgroups of driving faults examined in Table 19. Total faults was the best or equal-best predictor for three out of the four accident types examined in Table 19 and in the fourth type (male drivers’ non low-speed-manoeuvring accidents) it was nearly as good as the best predictor (normal driving faults).

### 9.3.10 Factor and discriminant analyses of driving faults

The sub-categories of faults used in the modelling above were established by Maycock and Forsyth (1997) on the basis of judgement, rather than empirical analysis. It may, therefore, be possible to find better groupings to predict accidents and/or to represent underlying driver competencies or dimensions of the driving task. This was explored in two ways: factor analysis to attempt to categorise faults according to some underlying structure, and discriminant analysis to discover the best subset of faults for identifying accident-involved people. Neither approach produced meaningful or useful groupings of faults.

### 9.3.11 Characteristics of high-faults drivers

If a tightening of the failure criterion for driving faults were to be considered, it would be desirable to form a reasonably complete picture of the type of driver likely to fail the revised test, so that the likely effect of failing them can be judged.

DSA records show that women who pass the test tend to make more driving faults than men. It has been shown above that for men and for women, high faults drivers tend (a) to be slightly older than the others, (b) to drive lower mileages, and (c) to drive less often in the dark. It is not clear why this should be, though a speculative explanation suggests itself, involving poor real or perceived driving ability, low driving confidence and possibly restricted access to a non-school vehicle pre and/or post-test. This receives some support from self-ratings of relative speed and relative driving ability collected during the survey. High faults people tended to rate their driving as relatively...
slow, and relatively poor. Replies to other questions in the survey revealed that high faults people tended to have taken more driving tests than had low faults people, and to have had more hours of professional instruction, and fewer hours of pre-test driving supervised by a friend or relative (Table 20). As Groeger (2000) has suggested, people who find it difficult to learn to drive may gravitate towards professional instruction. One (speculative) mechanism here may be that their driving does not inspire ‘informal’ supervisors with confidence. Another may be that older learner drivers have greater difficulty in finding a supervising driver.

Table 20 Driving faults and learning to drive

<table>
<thead>
<tr>
<th>No. of driving faults</th>
<th>Hours of professional instruction</th>
<th>Hours with a friend or relative</th>
<th>Total no. of driving tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-6</td>
<td>28.9</td>
<td>25.2</td>
<td>1.73</td>
</tr>
<tr>
<td>7-11</td>
<td>33.6</td>
<td>21.6</td>
<td>1.86</td>
</tr>
<tr>
<td>12-13</td>
<td>34.6</td>
<td>21.1</td>
<td>1.92</td>
</tr>
<tr>
<td>14-15</td>
<td>35.1</td>
<td>20.1</td>
<td>1.95</td>
</tr>
<tr>
<td>Total</td>
<td>32.4</td>
<td>22.6</td>
<td>1.84</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-6</td>
<td>44.2</td>
<td>27.6</td>
<td>1.95</td>
</tr>
<tr>
<td>7-11</td>
<td>48.8</td>
<td>22.7</td>
<td>2.05</td>
</tr>
<tr>
<td>12-13</td>
<td>52.1</td>
<td>21.1</td>
<td>2.12</td>
</tr>
<tr>
<td>14-15</td>
<td>51.6</td>
<td>19.3</td>
<td>2.15</td>
</tr>
<tr>
<td>Total</td>
<td>48.3</td>
<td>23.6</td>
<td>2.04</td>
</tr>
</tbody>
</table>

9.4 Discussion

9.4.1 Does the survey support changing the failure criterion in the driving test?

The main purpose of this study was to discover whether driving faults during the driving test are predictive of accident risk in a way that would justify changing the current failure criterion of 16 or more driving faults. To do this, it was necessary to examine both the relation between driving faults and actual numbers of accidents, and to use statistical modelling to allow the influence of exposure to risk to be explored.

Statistical modelling showed that once the effects of mileage, age, and driving in the dark have been removed, reasonably strong predictive relationships between driving faults and accidents are revealed. This is true for ‘all accidents’ and for accidents not classified by the respondent as low speed manoeuvring accidents. 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 (ie 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 on its own 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. (Driving experience defined in the usual way – ie how long a person had been driving since the test – was the same for all participants since they were asked about their first six months of post-test driving.) 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. If high-faults people were to decide to drive more during their first six months of driving, we cannot be absolutely certain whether they would have more accidents (because of their increased exposure) or whether this would be counteracted by the fact that they would gain more experience.

9.4.2 Targeted vs. general safety interventions

The assumption underlying efforts to improve the ability of the driving test to identify drivers with a raised accident liability is that to do so will improve road safety. As observed above, in the period immediately following a change to the test failure criterion the test would identify more of these drivers and cause them to have more training and/or practice before driving unaccompanied. In the longer term, the new testing requirements would be expected to influence the training system such that more drivers would reach the new test standard before coming forward for test at all. While such changes are likely to be beneficial, there is no guarantee that bringing a group of relatively unsafe drivers up to a new test standard will also bring their accident liability down to that of drivers who would have met that standard anyway.

There is also the question of whether it is fair and cost-effective to attempt to target a group of relatively unsafe drivers. Possible reasons for such targeting include the following:

- A group whose standard of driving is relatively unsafe might be more susceptible to improvement by training and practice. For example, its members might have skill and knowledge deficiencies that could be remedied.
- A high accident group imposes higher costs on other members of society (including those who become involved in the group’s accidents).
- A group whose excess accidents are explained by high exposure may nevertheless offer more scope than a low accident group for reducing the road accident total.

The fact remains, however, that it may be possible to obtain greater or more cost-effective road safety benefits by improving training and practice for drivers in general,
or by tackling the wider aspects of the licensing system so as to reduce risk and control behaviour during the early months of post-test driving. Baughan and Simpson (2002) provide a review of some attempts to achieve such changes in other countries. A summary of the review is presented in Section 10 of this report.

9.5 Conclusions on tightening the failure criterion for driving faults

1. The results of statistical modelling indicate 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.

2. 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.

3. There is no test failure threshold for total driving faults that identifies a group of men who will have more accidents than the rest during the first six months of unsupervised driving. For women, there are cut-off points that identify higher accident groups.

4. The higher-faults, higher-accident group of women have fewer accidents than the men in any ‘number of faults’ group.

5. People who make a large number of less serious faults (driving faults) on test tend, at least during the first six months after the test, to be intrinsically less safe as drivers than people who make few faults. There is therefore a valid argument for tightening the current failure criterion (16 or more driving faults) to improve the ability of the test to identify potentially unsafe drivers, and to cause them to have more training and practice before driving unsupervised. However, high-faults drivers also tend to drive fewer miles, and drive less often in the dark, than low-faults people. This means that, at least during the first six months of post-test driving, there is no relation between number of faults and actual numbers of accidents for men, and only a rather weak one for women. The argument for tightening the failure criterion is therefore finely balanced: the change should have some road safety benefit, but might be considered by some to penalise unfairly the new group of people who would then fail the test.

6. The above conclusions are not changed if low speed manoeuvring accidents, or accidents for which the respondent feels little responsibility, are excluded from the analysis.

7. The survey showed that, in addition to the difference in mileage, age, and driving in the dark, high faults drivers tended to rate their driving as relatively slow, and relatively poor. They also tended to have more hours of professional instruction than low faults people, fewer hours of pre-test driving supervised by a friend or relative, and to have taken more driving tests than low faults people.

8. The survey provides no good evidence to support the setting of separate failure criteria for subcategories of driving fault.

10 Other ways of using the licensing system to improve safety

10.1 Introduction

To achieve some of the desired improvements in training and experience, the driving test may need to be supplemented by other measures designed to encourage a more structured approach to learning and the gaining of more pre-test experience.

It is also desirable to consider measures that control risk by changing exposure during early solo driving, or by influencing the way drivers choose to behave once they have passed their test. The aim is, or should be, for this to be done in ways that do not prevent drivers from gaining the experience necessary to reduce their accident liability – otherwise the effect could be simply to delay the excess accident liability of novice drivers until the time when unrestricted solo driving is allowed.

Revocation of licence for drivers who accumulate six penalty points within two years of passing the test is an example of such a mechanism in Britain. Examples of licensing measures for novice or learner drivers used in other countries include restrictions on night driving, carrying passengers, and vehicle power – and the setting of especially low alcohol limits. Periods of accident-free or conviction-free driving may be required before a driver can be fully licensed, or there may be a restriction on the number of penalty points a driver is allowed to accrue. There may also be requirements to undertake specified training. Other provisions include staged testing, staged training (better matched to the natural progression of skills and knowledge acquisition) a lengthening of the period of supervised learning, and other measures to encourage or require an increase in the levels of supervised practice.

Delays in full licensure to increase amounts of supervised practice also (a) takes advantage of the beneficial effect of age (maturity) on accident risk and (b) shortens driving careers, thereby further reducing total accidents.

Combinations of such measures are generally known as ‘graduated licensing systems’, a term reflecting the idea of a staged progression from initial learning to unrestricted solo driving. Such systems have been, and are being, introduced rapidly by licensing authorities in many parts of the world and where evaluations have been undertaken there is often evidence of positive effects.

The project included a review of some graduated and related licensing systems from other countries, including the results of evaluative studies where these were available. The full review is published as Baughan and
Simpson (2002). An overview of the main findings is presented here. No attempt has been made to update them for this report, but for an entry to the more recent literature the reader is referred to Hedlund and Compton (2005). The term graduated licensing system (GLS) is used here in its broadest sense. In some countries – eg North America – it tends to be reserved for systems involving on-road driving restrictions (Williams, 2000a).

10.2 Components of graduated licensing
This section summarises the results of studies reviewed by Baughan and Simpson (2002) that assessed the effects of the individual components of licensing systems and considers whether they might be suitable for introduction in Britain.

10.2.1 Extending the pre-solo learning phase
Several measures aim to increase the duration of the pre-solo phase of learning to drive: increasing the minimum age for solo driving, reducing the minimum age for learning, or specifying a minimum learning period. The main justification for extending the pre-solo phase is that it should increase the amount, and possibly the quality, of training and practice that a learner accumulates before being allowed to drive unsupervised. In addition, an extended period can provide a framework that allows a structured training programme to be introduced, as is being planned in Sweden and Norway, for example.

There are several reasons to expect benefits from increasing the amount of practice accumulated by L drivers. The obvious one is the well-documented decrease in accident liability that currently occurs as experience is gained during the first few years of unaccompanied driving. This is probably associated with improvements to the novice driver’s control skills and rules/knowledge base; and to the development of higher order skills such as those associated with detecting, interpreting and reacting to hazards – skills that rely on the driver having developed good mental models of the traffic system. Clearly, such skills and knowledge need to be based on repeated experience of a variety of driving conditions and situations, and it is also likely that they can best be achieved after vehicle control skills have become sufficiently automated to free-up attentional capacity.

If the extension to the learning period is achieved by increasing the minimum age for solo driving, there are other potential benefits to road safety in that (a) drivers are more mature when they first drive solo and (b) driving careers are shortened, at least for people who start to drive as soon as they are permitted. Potential disbenefits of any measures that increase the licensing age include reduced mobility, and increased illegal unlicensed driving. There may also be a shift to transport modes that are less safe, if these are available to people at an earlier age; but the shift may instead be towards public transport, which would improve safety.

Increasing the gap between minimum age for starting to learn to drive and minimum age for solo driving may have its effectiveness dulled because (a) people may choose not to start learning until they are near the solo licensing age and (b) the measure does not affect people who start learning to drive later in life. An approach that aims to overcome these problems is to specify a minimum period between starting to learn and taking the test – in other words a minimum period for which a learner’s permit (provisional licence in the UK) has to be held before test. This has been done, for example, in California, Florida, North Carolina, New Zealand, Australia (Victoria), Nova Scotia and Ontario. Following a review of graduated licensing in Australia, Triggs and Smith (1996) suggested that such an approach is more effective (at ensuring that a minimum amount of experience is gained before licensing) than simply having a lower minimum age for a learner’s permit and higher minimum age for a full licence.

All measures that extend pre-solo experience are likely to lead to an increase in the number of accidents during learning – though accident rates during supervised driving are generally very low so the increase in numbers would be small. The first Cohort study (Forsyth et al., 1995) estimated that the accident rate per mile during the first year of driving was of the order of 18 to 36 times higher than the rate during learning. More recently, a study conducted in Britain for the EC project BASIC, found that 3.2% of a sample of over 10,000 people who passed the driving test reported having had an accident while learning to drive. These people reported a mean of around 62 hours of driving while learning to drive (Peraaho et al., 2004). Measures to extend pre-solo experience are also likely to increase the cost of learning to drive, though this may not be inevitable, given the cost of repeatedly failing the test under the current system. A simple reduction in the minimum permitted age for learning to drive may be accompanied by a tendency for people to obtain their full licence earlier than they do at present. This would tend to increase the total mileage driven in a driving career, and to reduce the maturity of novice drivers – both of these changes tending to increase accidents. Increasing the minimum age for solo driving, or introducing a minimum learning period may lead to an increase in illegal unsupervised driving, a shift to other, less safe transport modes and/or increased travel as car passengers with novice drivers. Such disbenefits may, of course, be far outweighed by the accompanying improvement in novice driver safety. Any shift towards public transport would provide additional safety benefits.

In Sweden, where the minimum age for solo driving was 18 yrs, it was relatively easy to introduce extended learning periods because this could be done solely by reducing the minimum age for starting to learn to drive from 17 yrs 6 mths to 16 yrs, thus providing young people with an opportunity, rather than a restriction. In Britain, where the minimum age for solo licensing is 17 yrs, following Sweden by introducing a two year period between minimum age for accompanied driving and minimum age for solo driving would involve an increases in the solo licensing age of at least a year, unless the minimum age for starting to learn to drive were to be reduced to below age 16. A shorter minimum learning period of six months or a year would be easier to
accommodate; though, for many drivers, setting a six month minimum period might not increase their actual period of learning. The current Cohort study being undertaken by TRL for DfT is showing that 82% of candidates already take longer than six months to pass their driving test, and that 45% take longer than a year. Also, 75% of learners do not take their first test until they have been learning for more than six months. 32% of learners do not take it until they have been learning for more than a year.

A related option is to specify the training and practice that should be accumulated before taking the test – this could be done in the form of a legislative requirement and/or by encouraging it as good practice, perhaps backed up with voluntary or mandatory logbooks. Maryland, California, Michigan, and Western Australia provide examples of systems that specify minimum levels of supervised practice at either the pre-solo or intermediate phases of learning to drive.

Sweden, Norway and Finland are examples of countries that specify some aspects of the training required during the pre-solo phase of learning to drive. The effectiveness of such measures has not yet been convincingly demonstrated, but there are strong arguments that they ought to be effective if the content of the training is optimised, and strong indications of what this content ought to be (see, for example, Siegrist, 1999; Hatakka et al. (2002)).

Changes in licensing ages, and other measures designed to increase the amount of supervised practice, have often been introduced as part of wider changes to a licensing system. However, this was not so in Sweden, which therefore provides good evidence of the link between supervised practice and safety.

When the Swedish minimum permitted age for learning to drive was reduced from 17.5 to 16 years, about half of the young novice drivers made use of the system and, in so doing, increased the amount of pre-test driving experience by a factor of 2.5 to 3. The effect, averaged over all the eligible young drivers, whether or not they actually used the new provision, was estimated to be a 15% drop in accidents per mile, and a slightly higher drop in accident liability (Gregersen, 1999).

In Norway, reducing the minimum age for learning to drive from 17 to 16 years (but keeping the licensing age at 18), had only minimal effects on the amount of pre-test driving experience (Sagberg, 2001a). Not surprisingly, there was no detectable effect on post-test accident risk. Clearly the contrast between Norway and Sweden needs to be understood before the effectiveness of such a measure in Britain can be predicted with any confidence. Unfortunately, the explanation is not at present known, and we can only speculate. There would appear to be several possibilities, including the following:

- The period between minimum age for starting on-road training and minimum licensing age was quadrupled (from 6 months to 24 months) in Sweden, but only doubled (from 12 months to 24 months) in Norway. The scope for increasing the amount of experience gained in Sweden was therefore effectively double that in Norway. It also seems possible that the old six-month period may have limited the amount of pre-test experience in Sweden much more than did the old 12 month period in Norway. In other words, learner drivers in Norway may have generally been able to satisfy their perceived need for pre-test experience before reaching the minimum age for taking the test under the old system, whereas this may not have been so in Sweden. Here it would be relevant to examine the actual amounts of pre-test experience in each country. Sagberg’s results are presented in terms of numbers of trips/lessons, and number of km driven, whereas Gregersen’s are in terms of hours of driving; but a rough comparison does not suggest that under the ‘old’ systems Swedish learners drove much less than did Norwegian learners.
- It might be that under the old systems the culture of informal, supervised practice for learner drivers was less well established in Norway than in Sweden – though this is not obvious from the data on amounts of pre-test experience mentioned above.
- When the age limit for learning was reduced in Norway, this was accompanied by a number of other changes that tended to liberalise the system of learning to drive – for example, dropping the requirement to obtain a recommendation from a driving school, and a reduction in mandatory training. It seems possible (though this is speculation) that learners and supervisors may have interpreted these other changes as indicating that they needed to put less effort than before into learning to drive.
- The tightening of the criteria for supervising drivers that took place when the minimum learning age was lowered will have reduced the supply of supervising drivers. In principle, this might have had more of an effect in Norway than in Sweden. For example, in Sweden, the minimum age for a supervising driver is 24 years, and in Norway 25 years. However, in Sweden, formal permission for driver training is required, and the accompanying driver must be registered, whereas no such registration is needed in Norway. This should tend to make informal supervision easier to arrange in Norway.
- The proportion of young people who obtain a driving licence is considerably higher in Norway than in Sweden. In 1997, less than 30% of 18 year olds held a licence in Sweden, whereas the corresponding figure in Norway was over 50%. Sagberg (2001b, personal communication) has suggested that there may be a stronger tendency in Sweden for young drivers to come from those socio-economic groups in which parents are more likely to provide resources for informally supervised driving.

In the US, increasing the minimum licensing age for drivers was found to be effective in reducing crashes among new drivers (Williams et al., 1983). As mentioned above, a possible problem with extending the learning period is that it may encourage younger drivers to drive unsupervised as well as supervised, and therefore increase the incidence of illegal, unlicensed driving. This problem was highlighted by Williams and Mayhew (1999) in their review of US licensing systems, some of which
allowed learning to start at age 14. The authors recommended that graduated licensing systems in the US should maintain or raise the minimum learning age to 16. It should be noted, though, that an increase in unlicensed driving would not necessarily outweigh the safety benefit of an increase in the learning period, and could presumably be tackled by improved enforcement measures.

Preusser et al. (1985) examined the effect of delaying full licensure on mobility and independence of young people. They surveyed the lifestyles of 16 year olds in seven US states where many, few or no 16 year olds were licensed. The results indicated that delays in full licensure did not significantly hinder social activities. However, the socio-economic consequences of delaying full licensure in Britain from age 17 to age 18, or specifying a minimum duration of 12 months for the learning period, may be greater. This is because travel to work, university or college might be expected to be more important for the age group affected.

Changing the licensing ages, or introducing a minimum learning period seems well worthy of consideration in Britain although they may require some amendments to the information shown on driving licences to enable date of issue of a licence to be checked. As discussed above, introducing a minimum learning period has several advantages over changing the licensing age. The Environment, Transport and Regional Affairs Committee (1999) recommended a six or twelve month minimum learning period, and mandatory carrying of the driving licence. Mandatory or advisory prescriptions for the amount and type of supervised practice also need to be considered seriously. A learner driver’s logbook would provide a possible aid to implementation.

10.2.2 Night driving restrictions

Young drivers are over-represented in night-time single vehicle accidents. Night driving restrictions can be used to postpone driving at night – at least for high-risk recreational journeys – until novice drivers have more experience of daytime driving. Hampson (1989) argued that this type of restriction would mirror restrictions already put in place by parents in requiring children to be home by a certain time. It would also serve an educational purpose to highlight the risk of late night driving to new drivers. Night driving restrictions have been widely introduced, especially in North America where Williams (2000b) reported that they had by that time been adopted by 25 states.

A number of studies have evaluated the effects of night driving restrictions – see Williams and Preusser (1997) for a review – and the weight of evidence is that they are effective in reducing accidents during the restricted hours. Clearly the level of compliance will depend on a number of factors, and is to some extent open to control by the licensing authority. In fact, studies of several licensing systems have found 40-50% of drivers reporting that they had violated the curfew restrictions, though the level of compliance is generally reasonably good.

A potential problem with curfews is that people may change their journey patterns to avoid the curfew hours, such that accidents increase during the time leading up to the start of the curfew. In Pennsylvania and Maryland where the curfew did not start until midnight or 1am respectively, an increase in accidents amongst the ‘curfew-affected’ drivers in the preceding 2-3 hours was reported, this being offset by a continuation in accident reduction in the first few hours after the curfew ended. However, the general weight of evidence appears to be that night driving restrictions do not produce important increases in accidents outside the restricted hours (Williams, 2000a).

To the extent that night-time restrictions are observed, it is not surprising that accidents during the curfew periods decline. This effect in itself will benefit road safety if it is not outweighed by safety problems arising from secondary effects of the shift in driving patterns. The restriction saves night-time accidents by removing some night driving exposure from a driver’s driving career. However, the question arises as to whether night-time restrictions have the intended benefit of reducing the novice driver’s excess night-time accident liability once he or she is permitted to drive at night. Do the restrictions merely delay the problem that novice drivers face with night driving, or does learning during the restricted period mean that drivers are better able to cope once the restriction is lifted? There appears to be little research evidence on such questions.

Some potential disadvantages of night-time restrictions are to do with effects on mobility and freedom. It might be supposed that such effects would make the measures unpopular, and that Governments would therefore be reluctant to introduce them. Resistance in Australia to night-time driving restrictions appears to have been associated mainly with effects on travel to employment and study, especially in rural areas. Surveys suggested that 82% of the population did not agree with late night driving restrictions for young drivers (AGB: McNair 1988). It was suggested that this could be overcome by having a restriction that starts sufficiently late, and allows exemptions for workers or in rural areas. This would lead to some reduction in the potential benefits but the risk is mainly associated with social driving. Exemptions for essential journeys have been included in most graduated licensing systems, including the system in Nova Scotia which is supported by nearly 40% of parents and around two thirds of teenagers in the system (Mayhew et al., 1998).

A survey of US teenagers’ awareness and attitudes towards night-time restrictions concluded that in states where night driving curfews existed, the majority of teenagers were in favour of them (Opinion Research Corporation, 1985). Another survey of teenagers (Williams et al., 1985) found that most students knew about and reported complying with night curfews in the states that had them. Williams and Lund (1986) reported that 73% of a nationally representative sample of parents with teenage children were in favour of night time restrictions. Waller and colleagues’ evaluation of the Michigan system also shows high levels of support for a system that includes a curfew for unsupervised driving between midnight and 5 am. (Waller et al., 2000). The restriction applies to people with a ‘Level 2’ licence – ie they must be aged at least 16, and have accumulated at least 50 hours of supervised driving experience (including a minimum of 10 hours at
be taken seriously. Parental involvement is included as part of the licensing systems in, for example, Michigan, Maryland and California, where a parent or another responsible adult must sign to confirm the number of hours practice that a learner has completed. In Britain also, the voluntary logbook introduced by DSA aims to encourage a structured approach to learning to drive, and to emphasise the importance of accumulating practice. As Waller et al. (2000) point out, such provisions are a way of seeking to ensure that learners do not respond to other licensing provisions simply by delaying licence – an action that would reduce accidents, but that would not have the desired effect of increasing practice under relatively safe conditions.

Parents, other supervisors, and learner drivers themselves have shown strong support for graduated licensing in general, and certified supervised practice in particular (eg Ferguson and Williams, 1996; Mayhew et al., 1998; Mayhew et al., 1999b; Begg et al., 1995; Waller et al., 2000). The question of how to encourage parent participation without penalising those whose parents will not, or cannot, help needs to be addressed. Whines (1988) reported that learners in New Zealand considered that finding a supervisor to allow them to drive during curfew hours or with passengers was inconvenient.

Studies of the French ‘Apprentissage’ system (Chatenet and Leroux, 1999) suggest that the effectiveness of parental supervision may depend on whether parents are motivated primarily by considerations of safety, or by economic aspects such as insurance discounts. Sagberg (2001b, personal communication) has also suggested that socio-economic factors might help explain why lowering the minimum learning age had very different effects on pre-test experience in Norway and Sweden. One implication here is that to maximise the effectiveness of parental supervision, attention needs to be given to helping and encouraging parent supervisors, especially those who are not themselves primarily motivated by safety.

In summary, increasing the amount of driving accumulated while learning to drive has generally been shown to produce valuable safety benefits, and ‘informal’ practice, supervised by a parent or other responsible adult, can play an important part in this. In Britain, any change designed to lengthen the period of learning to drive would be expected to increase supervised practice. However, the size of the increase cannot at present be predicted, as the contrast between Norway and Sweden discussed in Section 10.2.1 illustrates.

A logbook system would appear to be potentially useful in helping to structure the learning process and communicate to supervisors and learners the amount and type of practice that is needed. If a requirement for a minimum level of practice were to be introduced, this would presumably have to be certified by the supervisor. Such a requirement might be seen as an unwelcome burden on supervisors and, as such, difficult to introduce and vulnerable to dishonesty. However, the weight of evidence from other countries is of parents strongly supporting this and other provisions of graduated systems. Before introducing a requirement for certified practice in Britain the question of how to deal with learner drivers without easy access to a suitable informal supervisor needs to be addressed.

10.2.3 Informal supervision by parents or other adults

Informal supervision by someone other than a professional driving instructor is a feature of many licensing systems. Swedish experience (Section 10.2.1) shows that when pre-solo driving experience was increased substantially, post-test accident risk reduced. Informal supervised practice played a large part in this.

Several reviews (eg Brown, Groeger and Biehl, 1987; Horneman, 1993; Mayhew and Simpson, 1996; Williams, 1997) have concluded that professional tuition has yet to demonstrate benefits over private practice, though the first Cohort study (Forsyth, 1992) found that candidates who took no professional instruction at all tended to have a lower pass rate than other drivers. A study commissioned by DfT (Groeger and Brady, 2004) observed the rate at which learner drivers skills increased and found that this depended on the total amount of driving done rather than on the amount of informally supervised practice or the amount of professional training.

Some countries (eg Sweden, France and some North American States) require the supervisor to enter into an agreement with the learner driver and the licensing authority. This provides an opportunity to give the supervisor some basic information on what is expected and emphasises that supervising a learner driver is something to be taken seriously. Parental involvement is included as part
10.2.4 Passenger bans/restrictions

The problems associated with the carrying of passengers by novice drivers have been widely reported. Since young passengers are often carried by young novice drivers they, too, suffer from the novice drivers’ high accident liability. In addition, it has been shown that the presence of young passengers can adversely influence the behaviour of young drivers. In the USA it has been estimated that two-thirds of the deaths of teenagers as passengers occur in vehicles driven by teenagers, and 16 year olds contribute disproportionately to these deaths (Williams and Wells, 1995). A study of Ontario data (Doherty et al., 1998) found that for 16-19 year old drivers (but not for drivers older than this) accident involvement rates were about twice as high with passengers as without. Rates were also significantly higher for two or more passengers than with a single passenger. The effect was pronounced for both male and female drivers. Preusser et al. (1998) also found that passenger presence was associated with proportionately more ‘at fault’ fatal accidents for drivers aged up to 24, was a neutral factor for drivers aged 25-29, and was associated with fewer ‘at fault’ fatal accidents for drivers aged 30 or over. Again, fatal accident involvement was especially high for teenage drivers with more than one passenger. Chen et al. (2000) also concluded that, for young drivers, accident involvement increased if they were accompanied by passengers of similar age.

Baxter et al. (1990) examined the influence of passengers on driver behaviour and found that signalling before changing lanes was reduced in the presence of younger male or older female passengers. The tentative explanation was that signalling may be a relatively peripheral task which is sacrificed when drivers attend to conversations with passengers, and that younger male and older female passengers are perceived to have extreme attitudes towards driving violations, thus demanding more attention as the driver tries to match behaviour to the perceived standards of the passenger. Drivers accompanied by older female passengers tended to drive more slowly than drivers with younger passengers or no passengers. There was a non-significant tendency for drivers with younger male passengers to drive faster than those with no passengers. The study could not show whether the passenger actually influences speed, or whether people who carry older passengers are in some way different from other drivers. However, the authors speculated that drivers may be unwilling to exhibit irresponsible, high risk, behaviour in the presence of an older person. They argued that the influence of passengers on speeding behaviour can best be understood in terms of the driver adjusting his behaviour to conform with what he believes to be the passenger’s norm for ‘good’ driving. Rolls et al. (1991) reported that young males considered that their driving style was adversely affected by the presence of friends as passengers but was positively affected by the presence of their parents or their girlfriend/spouse. Young drivers of both sexes assessed journeys with friends as slightly more risky than journeys with their partner or spouse. A more detailed follow-up study amongst male drivers (Rolls and Ingham, 1992) confirmed that drivers tended to adopt different driving styles depending on the type of passenger. They suggested that drivers had two main reasons for this: that passengers expected them to drive in a certain way and that they felt a greater responsibility when driving. Those male drivers classed as ‘safe’ tended to say they drove least safely alone and more safely with passengers regardless of their type. ‘Unsafe’ male drivers drove least safely alone and also when accompanied by their male friends. Waylen and McKenna (2000) found that, of drivers with an estimated age of 17-24 years, males with a male passenger, and females with a male passenger, drove faster than lone drivers. The speed of young female drivers was not associated with the presence or absence of female passengers. Male drivers with female passengers drove more slowly than lone drivers. The authors argued that this last result implies that to restrict young male drivers from carrying female passengers would be to deny them the safety benefit of the reduced speeds.

Many graduated licensing systems have restrictions on the carrying of passengers. In Nova Scotia, no passengers other than the supervising driver are allowed at the learner phase. At the newly licenced phase, only one front seat passenger is permitted, and only as many passengers in the rear seats as there are seatbelts. In Ontario, only the supervising driver is allowed in the front seat although other passengers can be carried; in the newly licenced phase the only restriction on passengers is the number of seat belts. Indiana and Massachusetts have restrictions on passengers during the learning phase. In California, no passengers under 20 are allowed in the first 6 months of the intermediate phase unless accompanied by a driver aged 25 or more. This restriction does not apply once the driver reaches the age of 18, and is relaxed in certain circumstances – eg taking younger siblings to school. In Georgia no more than three passengers under 21 are allowed during the intermediate phase (unless they are family members), and in Delaware a maximum of 2 passengers is allowed in the second 6-month period after the 6-month learner phase. In New Zealand, drivers at the ‘restricted licence’ phase may not carry passengers unless a fully licensed adult is present. This appears to have reduced teenage passenger injuries in vehicles driven by other teenagers (Frith and Perkins, 1992).

Given the association between passengers and accidents, particularly amongst teenage drivers, restricting passengers for drivers when they first begin driving unsupervised is an option that merits serious consideration. In Britain, there is currently no requirement for drivers/passengers to carry evidence of age or, indeed, a driving licence. This may need to be changed to facilitate enforcement though, for many young drivers, their parents will provide effective enforcement. The possibility of young people making more car journeys as car drivers rather than passengers, thus increasing the total number of journeys, would need to be taken into account before a decision were made on whether to introduce passenger restrictions. So, too, would social effects such as reduced mobility, and impact on schemes to reduce drink-driving by promoting the idea of ‘nominated drivers’. 

57
10.2.5 Reducing alcohol limits

An International Symposium on ‘Young Drivers Impaired by Alcohol and Other Drugs’ (Benjamin, 1986) concluded that for a substantial proportion of young drivers, accident risk increases at lower concentrations of alcohol than is the case with older and more experienced drivers. This confirmed the earlier findings of the Grand Rapids study (Borkenstein et al., 1974) which showed novice drivers and inexperienced drinkers to be at increased risk.

A review of drinking and driving in Great Britain (Maycock, 1997a), showed that drivers in the 20-24 year age group were the most likely to be found over the limit following an accident and also had the highest number of drink drive accidents relative to injury accidents in general. The incidence of drink driving was somewhat less for 16-19 year olds but their risk in terms of the number of drink drive accidents per 1000 injury accidents was close to that of the 20-24 year olds.

Evaluations of those countries that have included lower alcohol limits for novice drivers as part of their licensing systems generally show that the limits appear to be effective in reducing accidents (Benjamin 1986). Research by Hingson et al. (1989) also showed that lowering the BAC limit for young drivers was effective in reducing night-time fatal crashes among teenagers in Maine. In Ontario the graduated licensing system introduced in 1994 included a zero BAC requirement for novice drivers. This was accompanied by an overall reduction of 27% in the incidence of collisions involving alcohol use for novice drivers (Boase and Tasca, 1998); although it is not clear how much of this reduction was due to some of the post-GLS sample driving under supervision.

Given the problems of alcohol related accidents amongst novice drivers, imposing lower limits on young or novice drivers is likely to bring benefits and may also instil safer drink/driving habits even after the restricted period ends. However, in Great Britain, enforcement of a differential BAC limit for novice drivers would be difficult in the absence of a requirement to carry licences or identity cards. The view has sometimes been expressed that it may be counterproductive to have a lower limit for novices, who would then see the limit ‘raised’ just as they moved into the group in which the drink driving problem peaks.

10.2.6 Increased consequences of traffic violations

The association between violating behaviour (ie departure from the formal or informal rules of good, safe driving) and accident liability is well documented. The nature of this link remains unclear, so the extent to which the increase in accidents happens as a direct result of rule-breaking behaviour is not known. This means that the extent to which accident liability would be reduced if the rule-breaking behaviour were eliminated is not known either. Nevertheless it seems highly likely that rule breaking behaviours do have a direct influence on accident liability. They increase opportunities for conflict with other vehicles, reduce the time available to deal with hazards, increase mental workload, put increasing demands on vehicle control skills, reduce predictability to other road users and make the consequences of error more severe.

Encouraging novice drivers to comply with traffic rules is also likely to be beneficial in other respects since, for example, drivers who habitually break the rules are effectively putting themselves out of the reach of future rules that may be introduced to promote safety.

Many licensing systems contain elements that require a period of conviction-free driving before moving to the next stage, or enable licence sanctions to be introduced at a lower threshold than is the case for fully licensed drivers. The British system of licence revocation for drivers who accumulate 6 penalty points within two years of passing the driving test is an example of such a restriction. Measures like this can be seen as a way of maintaining a supervisory influence on novice drivers during their period of early solo driving and are attractive in that they seek to address motivational aspects of the driver safety problem.

McKnight (1996) suggested that licence sanctions can reduce exposure for a sub-group of new drivers who are inclined towards unsafe driving, but can also encourage safer driving by acting as a deterrent. Haworth et al. (1995) argued that motivation to drive more safely can be increased by making the quality of the driving record the determining factor in removing restrictions or imposing sanctions rather than just including time-based restrictions. There is, as yet, little evidence on effectiveness of such measures. However, in New Zealand, self-reported violations fell after graduated licensing was introduced, suggesting that novice drivers became more conscious of the law, or that the sanctions during the probationary period were sufficient to deter them from violating (Whines, 1988).

The adequacy of the legislation and enforcement, and the severity of the penalties applied, are likely to be major factors in the success of any system that seeks to impose on novice drivers special penalties for traffic violations. However, penalties that delay a driver’s exit from a graduated licensing system, or prolong a probationary period, run the risk of encouraging drivers to violate the licensing system itself by driving outside the provisions of the licence. This may be happening in Britain, since drivers who accumulate six penalty points within two years of passing their test have a rather low relicensing rate. Simpson et al. (2002) found that just over half of them had not regained a full licence two years after the penalty, though subsequently there were signs that the relicensing rate was increasing). DfT- sponsored research on unlicensed driving (Knox et al., 2003) found that about 19% of revokers who responded to a survey admitted to unlicensed driving.

10.2.7 Increasing quality or quantity of education/training

Licensing systems may, in principle, seek to improve or increase driver training by a number of mechanisms – for example:

- Requiring specified training at one or more phases of licensing.
- Increasing the duration of the learning period.
- Introducing new testing requirements.
- Allowing a swifter exit from a licensing stage for drivers who take specified training.
Unfortunately, the benefits of driver training are at present unclear. For example, Mayhew & Simpson’s (1996) review found little evidence of a relationship between driver education/training and accident involvement. The authors concluded that, until more work has been done on ways to make training more effective and/or to strengthen the relation between training and accident involvement, they could not recommend introducing training as part of graduated licensing, or allowing accelerated progress through the system for people who take optional training. Reflecting such concerns, the time discount allowed for training in New Zealand was reduced in 1999.

There are many possible reasons for the apparent lack of effectiveness of driver training, and these have been discussed extensively in the literature. A number of themes have emerged, which offer the hope of improving the effectiveness of training. For example, current training of learner drivers tends to concentrate on car-control skills. This may increase drivers’ confidence in their ability to cope in difficult situations, and neglect higher order skills associated with hazard detection and evaluation, and decision making, which was one of the reasons for introducing the hazard perception test into the theory test in November 1999. Factors associated with the attitudes, goals and motivations of drivers also appear to be important influences on driver behaviour, but are not well covered in most conventional training. Thinking on these issues was developed as part of the recent EU project, ‘GADGET’ (Hatakka et al., 1999; Hatakka et al., 2002). Training that addresses these deficiencies may well prove to be effective in improving safety. Developing and evaluating such training is an important task. It will be desirable to support this training by finding ways of assessing the relevant knowledge and skills in the driving test – an area that has been considered as part of the EC project TEST (Baughan et al., 2005).

10.2.8 Restrictions on speed or on using high speed roads

Some graduated licensing systems have incorporated speed restrictions (eg for learners in France) or restrictions on the use of higher speed roads (eg for learners in the first stage of the Ontario system). In Northern Ireland, a maximum speed limit of 45 mph applies before the test and during the first year after passing it.

McKnight (1996) suggested that placing lower speed limits on new drivers is of questionable benefit as the speed differential may be more hazardous than uniformly higher speeds. Doherty and Andrey (1997) estimated that restricting new drivers from driving on high speed roadways (ie equivalent to motorways in Britain) is likely to increase accident involvement as these are some of the safest roads. Depending on how the restriction is framed, it could also prevent testing and training at higher speeds, and delay the gaining of experience – indeed in Northern Ireland some consideration has been given to possible removal of the speed restriction on new drivers for just such reasons.

10.3 Conclusions on licensing provisions

10.3.1 The case for graduated licensing in Britain

The review identified several elements of licensing systems that have been effective in other countries, address recognised aspects of the novice driver safety problem, and could be considered for introduction in Britain. However, even where there is good evidence that an element has been effective in another country, prediction of the likely benefits in Britain is difficult. The main reason is that the current situation in Britain is rather different from the situation that existed in other countries before their graduated licensing systems were introduced. In particular, New Zealand, and many States in North America, permitted driving to start at age 15 or below, with minimal requirements for supervision, training and testing; and few, if any, licence restrictions. Some States still permit unsupervised driving from age 15.

Despite this caveat, a serious case can be made for introducing some elements of graduated licensing, or graduated learning, systems in Britain. First, apart from the difference in licensing age, for some drivers the current British system is perhaps not so different from the pre-graduated licensing situation in the North America. Those who are able to develop vehicle control skills and observational procedures quickly, can find themselves driving unsupervised and unrestricted after very small amounts of training or practice.

Secondly, results from Sweden indicate that increasing the amount of experience gained by learner drivers while they are being supervised by another driver is very effective at reducing their accident liability once they are allowed to drive unsupervised. In other words, it appears that some of the learning responsible for the steep decline in accident liability currently seen in the first year or two of solo driving in Britain would take place in relative safety if the driver were being supervised.

Thirdly, novice drivers in Britain do have problems with alcohol, night driving, and passengers, and there is good reason to expect benefits from measures that address these problems directly.

Fourthly, the emerging indications of where current driver training and education are deficient, and how they could be improved, offer the possibility of our being able to develop training that could be incorporated in licensing requirements with reasonable confidence that it would improve safety.

10.3.2 Elements likely to improve road safety in Britain

A review of published evaluations of licensing systems in other countries and of some other relevant research, indicates that the following elements would be likely to have beneficial effects if they could be introduced in Britain:

*Increasing the amount of driving experience accumulated by learner drivers.*

Increasing the amount of driving experience accumulated before solo driving has generally been shown to reduce novice drivers’ accident risk. If such increases in experience could be achieved in Britain, there would
almost certainly be an improvement in novice driver safety. Possible mechanisms for achieving this increase in experience include increasing the minimum age for holding a full licence, reducing the minimum age for starting to learn to drive on the road, and/or introducing a minimum learning period. Increases in pre-solo experience can also be achieved by specifying suitably high minimum amounts, and requiring supervisors to certify that the requirement has been met. Indications from the USA are that there could be a high level of support for this from parents and learner drivers. Advisory minimum targets for pre-solo experience also seem likely to be helpful. The introduction of a logbook for learner drivers should be useful in this regard.

Potential disbenefits of increasing the minimum licensing age include a reduction in the mobility and independence of young people, and a shift to other transport modes, some of which may be less safe; these would need to be considered before a decision were made. Reducing the age for starting to learn to drive without changing the licensing age for solo driving is likely to increase the number of accidents during learning – though accident rates during learning are very low, so the increase in numbers of accidents would be small. There would probably be a tendency for people to obtain their full licences somewhat earlier than they do at present. This would tend to increase the total mileage driven in a driving career, and to reduce the maturity of novice drivers – both of these changes tending to increase accidents. The size of such effects is not known at present.

Sweden achieved large increases in the amount of pre-test experience, and substantial improvements in novice driver safety, from reducing the minimum age of learning to drive from 17.5 to 16 years – so this type of measure merits very serious consideration in Britain. Unfortunately, results from Norway have been less encouraging, with small changes in pre-test experience, and no observable change in accident risk, accompanying a reduction in minimum learning age from 17 to 16 years. The reasons for these differences need to be better understood if a reduction in learning age is to be considered for Britain. Evidence on the shortcomings of the Apprentissage system in France would also need to be taken into account.

Night-time restrictions

There is evidence that these can be very effective at reducing night-time accidents during the months covered by the restricted licence – though clearly this will depend on the level of enforcement and penalties. There have been indications of increases in accidents during the pre-curfew hours compensated for by reductions in accidents after the curfew hours but the weight of evidence is that curfews do not generally lead to important increases in accidents outside the curfew hours. If night-time restrictions were to be considered in Britain, the likely benefits would need to be compared with the effects on employment and mobility. Many of the curfew restrictions in other countries apply to drivers below age 17, whereas in Britain restrictions imposed during a probationary period would apply to drivers of 17 years or upwards who may be more likely to be using their cars for access to work or education. Nevertheless, the impact of the restrictions on mobility could be minimised by careful choice of the curfew period and by exemptions for work-related and other essential journeys, as is commonplace with curfew systems. Enforcement, and the associated need for carriage of driving licences or other forms of identification, would also need to be considered, especially as a reasonable proportion of the drivers affected will have moved away from the supervisory influence of the parental home.

In some jurisdictions curfews apply for pre-solo driving only. Such provisions are probably not desirable in Britain because they would restrict, rather than broaden, the experience gained during supervised driving.

Passenger restrictions

Given the association between passengers and accidents, particularly amongst teenage drivers, passenger restrictions for drivers when they first begin driving unsupervised is an option that should be considered. In Britain, there is currently no requirement for drivers/passengers to carry evidence of age or, indeed, a driving licence. To facilitate enforcement, this would need to change, though parental influence may be sufficient for some drivers. The possibility of young people making more car journeys as car drivers rather than passengers, thus increasing the total number of journeys, would need to be taken into account before a decision were made on whether to introduce passenger restrictions. So, too, would social effects such as reduced mobility, and impact on schemes to reduce drink-driving by promoting the idea of ‘nominated drivers’.

Increasing penalties for traffic violations

The association between traffic violations and accident liability, as well as other considerations, suggest that reducing novice drivers’ propensity to commit traffic violations would be beneficial to safety. Many licensing systems contain elements that require a period of conviction-free driving before moving to the next stage, or enable licence sanctions to be introduced at a lower threshold than is the case for fully licensed drivers. The British system of licence revocation for drivers who accumulate six penalty points within two years of passing their driving test is an example of such a provision. Such measures can be seen as a way of maintaining a supervisory influence on novice drivers during their period of early solo driving. There is, as yet, little evidence on their effectiveness, but they are relatively simple to introduce and are attractive in that they seek to address the motivational components of novice driver safety.

The adequacy of the legislation and enforcement, and the severity of the penalties applied, are likely to be major factors in the success of any system that seeks to impose on novice drivers special penalties for traffic violations. However, penalties that delay a driver’s exit from a

---

20 DSA introduced a voluntary logbook scheme for learner drivers in 1997. The revised version (the Drivers Record) has been sent out free with all new provisional car licences since June 2003.
graduated licensing system, or prolong a probationary period, run the risk of encouraging drivers to violate the licensing system itself by driving outside the provisions of the licence. Measures to combat this would need to be considered. Also, it is essential that the penalties are well-publicised, and used properly by the courts, if they are to act as deterrents.

**Improving training and education**

There are many possible reasons for the general lack of evidence that increased driver training improves safety, and these have been discussed extensively in the literature. A number of themes have emerged, which offer the hope of improving the effectiveness of training. For example, current training of learner drivers tends to concentrate on car-control skills. This may increase drivers’ confidence in their ability to cope in difficult situations, and neglect higher order skills associated with hazard perception and decision making. Factors associated with the attitudes, goals and motivations of drivers also appear to be important influences on driver behaviour, but are not well addressed in most conventional training. Thinking on these issues was developed as part of the recent EU project, ‘GADGET’ (Hatakka et al., 1999; Hatakka et al., 2002). Training that addresses these deficiencies may well prove to be effective in improving safety. Developing and evaluating such training is now an important research task. It will be desirable to support this training by finding ways of assessing the relevant knowledge and skills in the driving test (Baughan and Keskinen, 2005).

**Reduced alcohol limits for novice drivers**

Given the problems of alcohol related accidents amongst novice drivers, imposing lower limits on young or novice drivers is likely to bring benefits and may also instil safer drink/driving habits even after the restricted period ends. However, in Britain, enforcement of a differential limit for novice drivers would be difficult in the absence of a requirement to carry licences or identity cards. Also, drink-driving is more prevalent amongst those in their early 20s than it is amongst teenage drivers. It may therefore be counterproductive to have a lower limit for novices, who would then see the limit raised just as they moved into the group in which the drink-driving problem peaks.

**Driver and passenger identification**

Enforcement of several of the measures listed above would be difficult unless novice drivers, and possibly their passengers, are required to carry identification.

**Probationary licences and exit tests**

The British licensing system already includes a two-year probationary period after the practical driving test has been passed. At present, drivers who reach the end of the period without accumulating six or more penalty points automatically exit from the probationary phase. In principle, it would be possible to make this conditional on passing a further test, as is done in several graduated licensing systems. Such ‘exit tests’ may provide a suitable platform for assessing higher order skills and possibly habitual behaviours. In effect, the purpose of the exit test is to influence the training and/or practice accumulated prior to test, so that the desired skills are developed and undesired habitual behaviours do not become established.

There is not sufficient evidence available to allow the benefits of exit tests to be assessed at present, but it is possible to offer some general observations on their likely applicability to Britain. Further information may become available from a review of ‘advanced’ tests used in graduated licensing, commissioned by the Ontario Ministry of Transportation.

In Britain, current provisions during the probationary phase are relatively light. This means, first, that novice drivers are able to make full use of a car without being restricted by the need for supervision, passenger bans, or curfews. They are able to adapt their lives to the full benefits of car travel. Secondly, the provisions are demonstrably not sufficient to remove the novices’ excess accident liability during the probationary period.

An exit test introduced after two years of unaccompanied driving would need to induce significant changes in driver behaviour during that two year period if it were to have an impact on the excess accident liability of novice drivers, since much of that excess occurs during the first two years of solo driving. Indeed, to induce such behaviour changes, there would have to be important consequences for drivers who failed the test. Extending their probationary period would probably not be sufficient unless the probationary restrictions were significantly more severe than they are at present, since current restrictions allow full use of the vehicle. In principle, people who fail the exit test could be required to revert to accompanied driver status but, unless tests of very high validity and reliability could be developed, it would be hard to justify this after lifestyles had adapted to two years of solo driving.

In summary, simply adding an exit test to the end of the current two-year probationary period does not have much to recommend it in Britain. This conclusion might change if more severe, risk-reducing restrictions were to be imposed during the probationary phase since (a) drivers would not become accustomed to the benefits of unrestricted solo driving and (b) the current excess in novice driver accident liability would be controlled throughout the probationary phase. The function of the exit test would then be to ensure that, despite the restrictions in force, enough experience and learning occurred during the probationary period to mean that the novice driver accident excess did not simply transfer from the probationary period to the early years of unrestricted driving. Of course, it would be preferable to have as much of this learning as possible take place before the first driving test (ie the one that allows entry to solo driving).

One change that could well be useful would be to change the name of the post L-test licence to (say) the probationary licence. This should make novice drivers more aware of probationary conditions and facilitate enforcement of the types of restrictions discussed above. It would also provide a mechanism for making post-test
training or other provisions mandatory in the future, if and when the case for such measures becomes strong enough. Although it appears that EC legislation would not permit the introduction of a post-test probationary licence, it would be possible to issue the first full licence with a special code for a probationary period and issue a full ‘till age 70’ licence after this period (DTLR 2002)

11 The British practical driving test: summary of current status and possibilities for future development

11.1 Overall effectiveness

Given the relative levels of road safety in Britain and elsewhere, it appears that the driving test has proved rather effective – ie it has helped to achieve a reasonably competent and comparatively safe driving population. However, the fact that novice drivers’ accident liability is high immediately after the test but falls sharply as they gain post-test experience shows that they still have a lot to learn after passing the test. If it is possible for such learning to take place before drivers come for test – and evidence from other countries indicates that it is – there would appear to be scope to improve further the effectiveness of the test. That is, it would be beneficial if the testing system induced more of this learning to take place pre-test than it does at present.

There are four main ways of modifying a driving test to improve the competence and safety of novice drivers. These are: (i) broadening the test content to induce candidates to achieve competence in a wider range of driving tasks; (ii) improving the performance of the test as a measuring instrument, so that it is better able to distinguish between candidates who have, and have not, reached a desired level of competence; (iii) increasing the severity of the test, so as to improve the level of competence achieved; (iv) making more use of the test itself as a training opportunity (ie via examiner feedback). Though conceptually distinct, in practice these categories overlap since a change that addresses one of them may affect others.

The status of the British practical driving test, and the scope for improvement, are discussed below, as is the question of whether changing the test alone would be the best way of achieving the desired benefits.

11.2 Test content

The test provides reasonable coverage of most of the basic driving situations and skills; and most test centres and many routes have access to higher speed dual carriageway and national speed limit single carriageway roads. As a result the majority of learners will generally do some preparation on such roads even if their particular test does not include them. The accident data suggest that novice drivers have particular difficulties at night, and with judgement of appropriate speed (eg on the approach to bends and roundabouts in light traffic conditions). Ideally, therefore, the content of the test would be improved in these regards. However, speed selection on test is unlikely to be representative of the speed choices of unsupervised drivers. The difficulty of dealing with night driving in the driving test is obvious – and anyway, for novice drivers, night driving accidents may well be as much to do with factors associated with the recreational journeys being made as to problems of driving in darkness. It is therefore likely that the best solution lies in the training and licensing system, rather than within the test itself. Section 10 outlines some possibilities.

Novice drivers have relatively poor hazard perception skills. This situation should be improved by the recent introduction of a computer-based hazard perception test as part of the Theory Test that learner drivers have to pass before booking a practical test (Grayson and Sexton 2002). The practical test also assesses such skills: people fail if poor hazard perception or poor observational procedures lead them to make a serious or dangerous fault. It might be possible to strengthen further the assessment of hazard perception skills in the practical test. Commentary drives, or questioning the driver about a situation just experienced, would be worth exploring, though validity and practical feasibility would need to be assessed.

Strengthening this aspect of the practical test should encourage continued emphasis on hazard perception skills during practical training, and would require learners and their instructors to confront the difficulties of learning and using such skills at a time when basic control skills are occupying much of their attention.

There is growing realisation of the importance of drivers’ attitudes and goals as influences on behaviour, and the need for drivers to be aware of these influences and how to recognise and control them has implications for both driver training and testing. The practical test does not attempt to assess directly the attitudinal or motivational factors associated with the propensity of drivers to depart from good, safe driving practice once they begin to drive unaccompanied. However, it does not seem feasible to include such items in the pass/fail criterion because they are so open to manipulation by candidates. What might be useful, especially if combined with increased emphasis on examiner feedback, would be to include them in the practical test to emphasise their importance. There would be implications here for examiner training, and the current reluctance of instructors to participate in feedback sessions would need to be addressed – as would the receptiveness of candidates themselves to such feedback.

In fact, directly influencing driver training to emphasise the risks of such behaviour and to change attitudes towards it may be at least as valuable as modifying the driving test itself; and combining the two approaches seems likely to be the most effective option. Measures to penalise violational behaviour during early post-test driving could also be useful.

The EC project TEST (Baughan et al., 2005) has examined the possibilities for bringing these aspects into practical driving tests. That project, which involved audits of a large sample of driving tests, and surveys of examiners in six countries, also provided quantitative data on the test duration and content that examiners and
auditors believed is needed for a robust assessment of candidates’ driving.

A further aspect of driving not covered by the current test content is solo-driving decision-making, such as choosing where and how to conduct manoeuvres, dealing with missed turnings, route choice, and driving towards a specified destination by using direction signs. Given that novice drivers have to make such decisions immediately after passing the test, there is an argument for representing them in the test. The benefits of so doing, and the question of whether they would outweigh the difficulties and costs, require further examination. However, some other countries’ driving tests do cover such aspects, notably the ‘Practical Driving Assessment’ (PDA) of Drummond (2000) that has been adopted in Western Australia.

11.3 Administration of test technical standards

The Driving Standards Agency gives much attention to devising and selecting test routes, and to training examiners and monitoring their performance. Only people aged 25 and with extensive recent driving experience may become examiners. They have to pass a stringent driving test, lasting over an hour, followed by a panel interview. They then attend a residential four-week pass-fail course at DSA’s training centre. This features continuous assessments, progress checks, and final tests of theoretical knowledge, driving ability, and ability to conduct and assess a driving test to a consistent high standard. The probationary examiner is then attached to a driving test centre where his or her performance is further monitored and developed. Once in post, examiners are further monitored, and attend regular refresher training. Examiners’ written test reports are monitored regularly by the centre manager and frequently by DSA senior managers. Line managers regularly accompany each examiner on test, assessing and marking the candidate and then comparing notes with the examiner afterwards. Results of these observed tests show very good agreement between examiners on test outcome when they assess the same test. Senior examiners from DSA headquarters, independent from the examiner’s line management chain, regularly carry out audit inspections of test centres, in which they accompany examiners on test, and check documentation and procedures. Management information gives a regular analysis of the tests conducted by each examiner, and enables comparisons between examiners and between test centres. At least annually, managers carry out a detailed analysis of the tests conducted by each examiner, examining usage of test routes, choice of manoeuvres, number of faults of each type, and number of times the examiner has to take action to avoid a dangerous situation.

This is a very well-developed system for establishing and maintaining test standards. Nevertheless, there are variations in pass rate between routes, examiners, centres and regions – as shown in Section 5 of the report. These do not necessarily imply that there are undesirable variations in test standards, but a better understanding of the factors that affect pass-rates is desirable.

Clearly, differences in test routes within and between centres have a bearing on test standardisation. The report’s findings on the link between pass rate and free driving time, roundabouts and signal-controlled pedestrian crossings are relevant here, though since the survey was undertaken DSA has made some changes to test routes to improve coverage of road and traffic conditions.

Full route standardisation would be impossible to achieve because suitable routes are not available. Some within-centre variation between routes would anyway seem acceptable. Driving schools will tend to prepare pupils for the conditions encountered on all the routes used by a test centre, which are generally representative of the conditions in the local area. The test will therefore be fulfilling its main function of influencing training and practice. The fact that candidates are quasi randomly allocated to a test route will mean that there will be no systematic bias in favour of some candidates.

Between-centre differences in route difficulty cannot be justified in the same way, as they would mean that candidates in some parts of the country were being given a less challenging test than candidates in other areas. However, if routes are typical of conditions met by drivers in the area, the level of safety induced by the test in that area might be as good as the level of safety in areas with more difficult tests. A study of the extent to which drivers restrict their early driving to the area in which they take the test would be instructive in helping to assess the importance of regional differences in test conditions.

Between-centre differences in pass-rates merit further examination to discover whether they represent undesirable differences in route/traffic demands or marking standards, or whether they are explained by differences in the competence of candidates coming forward for test. Information on the amounts of training and practice accumulated by candidates at different centres or regions would be useful here. Recent work by DSA has found that test centres that have a high proportion of their candidates from socially deprived areas tend to have lower pass-rates than other centres, as do test centres in cities. These variables may be associated with poorer access to training and supervised practice.

If there are identifiable types of test centre at which candidates come for test particularly poorly prepared, then policies to address the problem may be needed. Clearly, low pass-rates at such centres would imply that the test is doing a useful job in screening out such candidates.

Further attempts to standardise test routes might involve observing driving only at certain features along the route. This approach has been taken in some Australian and North American states, but brings with it other potential problems. If such a change were to be considered for the British practical driving test, further evaluation would be needed.

The study identified a number of other factors that need to be taken into account when considering ways that the test could be improved. For example, the examiner selects

---

21 A theory test, including a hazard perception test, now forms part of the initial selection process for examiner training.
two out of three standard reversing manoeuvres for each test, and the probability of making a serious or dangerous fault associated with a reversing manoeuvre was found to be substantially lower for the turn in the road than it was for reverse parking, and reversing into a limited opening. The effect of the choice of reversing manoeuvres on pass rate is complex, involving interactions between difficulty of the manoeuvre, time of day of test, and possibly other variables. Nevertheless, the rather large differences between the manoeuvres in terms of the probability of making a serious or dangerous fault mean that this area merits further examination.

Drivers taking the test while it was raining had a pass rate about three percentage points lower than those taking the test in the dry. A further examination of the reasons for this might suggest ways of improving candidates’ ability to deal with such conditions.

11.4 The test marking system

The current marking system was generally liked by examiners and instructors who participated in the project. It achieves good agreement between examiners who observe the same test, and is ‘objective’ in the sense that it allows examiners to explain a failure decision by pointing to a specific event. It also represents a considerable investment by DSA in the training of examiners, and so is not something to be changed lightly. The system is, however, inherently limited in its ability to estimate a candidate’s underlying rates of making serious or dangerous faults. This tends to lead to low test–retest reliability, and may reduce candidates’ perceived control over test outcome, hence discouraging them from taking more training and practice before coming for test. In principle, it should be possible to improve the performance of the test as a measuring instrument by increasing the number of errors observed—i.e., increasing test duration or basing the test on less severe (more frequent) faults. However, the benefits of doing this are uncertain, and there would be practical difficulties. For example, as the report shows, a much longer test would be needed to make any substantial difference to its ability to estimate a candidate’s underlying frequency of making serious or dangerous errors. Unless a substantial reduction in pass rate (at least initially) were acceptable, it would become necessary to pass people who made a single serious or dangerous fault.

The fact that the test focuses on failures of performance rather than assessing good performance is also sometimes held to be a disadvantage.

Although the current marking system does have some limitations, what is not clear at present is (a) how important they are, and (b) whether alternative types of test are any better—i.e., whether they successfully solve the limitations without introducing new ones. Alternative types of marking system—e.g., based on examiner judgements of competence (as used in Sweden, the Netherlands and Western Australia), or based on the proportion of times a candidate performs an action correctly (as used in some Australian and North American states)—would need to be carefully evaluated before any decision on implementation was made.

Of course, any limitations of the current marking system would be reduced in importance if other measures were brought in to supplement the test as a way of influencing driver training and practice, as discussed in Section 10.

11.5. Predictive validity

It is impossible to measure the ability of a fully implemented driver licensing test to predict candidates’ future accidents, since people are not permitted to drive unsupervised until they have passed the test. If its ability to predict future accidents could be measured, there are many reasons discussed in this report to expect this to be rather low for real candidates, even for a test that is effective in maintaining safety standards. A second-order indication of predictive validity can be obtained by examining whether test passers who perform well during the test have fewer accidents than passers who perform relatively badly. The study described in Section 9 investigated the relation between driving faults (i.e., the less serious faults) on test and accidents during the first six months of post-test driving. The results indicate that once the effects of age, mileage driven, and driving in the dark have been adjusted for, there is a predictive relationship between accidents and number of driving faults. It therefore appears that high-faults candidates do tend to be intrinsically less safe as drivers than candidates who make few faults on test. In principle, the ability of the test to identify relatively unsafe drivers could be improved by tightening the failure criterion for driving faults. However, the argument for this is finely balanced since high-faults drivers tend to do less driving, and this weakens or removes the relation between test faults and actual accidents. Note that the survey was only able study the accident liabilities of people making less than 16 driving faults, since those making 16 or more failed the test.

11.6 Reliability

The between-examiner reliability of the test appears to be good, as evidenced by the results of routine check tests conducted by DSA. This form of reliability—i.e., the degree of agreement between pairs of examiners observing the same test—is the one most under DSA’s control via selection, training, and monitoring of examiners.

Differences between tests associated with routes, time of day, traffic levels, chance events and weather conditions will introduce another source of unreliability. The test marking system, in which a single serious or dangerous fault leads to failure, makes it vulnerable to such unavoidable between-test differences. Such effects will tend to reduce test–retest reliability—i.e., the level of agreement between pairs of tests conducted (for experimental purposes) on the same candidates.

A major influence on measured test–retest reliability is the fact that candidates choose when they come for test. Many appear to do so at a time when their own level of competence is such that they have only a moderate probability of passing. Their performance is not consistent: their underlying rate of making serious or dangerous faults means that they may or may not make one in any given
test. Again, this characteristic can be addressed to a certain extent in a driving test: the test can, in principle at least, be adjusted so that it is able to estimate more precisely a candidate’s underlying rate of making faults. Modifications to the test of the sort discussed in Section 6.5 could reduce these components of unreliability. However, even if the precision of the test were to be improved, candidates might still choose to come for test when they are in that range of competence within which the test is not able to discriminate well.

11.7 The test standard

The fact that candidates tend to come for test at a level of competence that gives them only a moderate probability of passing is of fundamental importance in understanding how the practical driving test operates. It means that, in terms of their underlying rate of making serious or dangerous faults, many people who pass the test are no better than those who fail. This is not necessarily a problem for road safety, and should not in itself be regarded as a deficiency of the driving test. What it does mean is that the test needs to be sufficiently demanding that candidates who have only a moderate probability of passing are nevertheless good enough to be permitted to move to the next stage of licensing. In Britain this involves being permitted to drive unsupervised with one probationary restriction.

If we judge that people with only a moderate probability of passing the test are not competent enough to drive unsupervised, then several types of solution are possible:

a Increase the severity of the test. One way of achieving this would be to increase its duration while keeping the failure criterion unchanged. A problem here is uncertainty as to how effective the new, more severe, test would be in inducing learners to reach higher standards. This is likely to depend on the degree to which learners are able to reduce their error rates by taking reasonable amounts of training and practice.

b Ensure that drivers who pass the test do not immediately go on to drive unsupervised and unrestricted.

c Persuade learner drivers to delay coming for test until their probability of passing, and therefore their competence, is higher. Pass-rates would increase, the standard of novice drivers would improve, the number of re-tests would decrease and test–retest reliability would increase. The degree to which this could be achieved in practice is not at present known.

In summary, it is possible to argue that the severity of the current test is too low, as evidenced by the current novice driver safety problem. However, driver competence and safety might best be improved by means of mechanisms other than raising the test standard. This is discussed further in the following section.

---

22 i.e. revocation of licence if 6 penalty points are accumulated within two years of passing the test.

11.8 Pass-rate and candidates’ competence

The pass rate in 2002/3 was 43-44%, and it appears that a substantial proportion of candidates have individual probabilities of passing of around this value. It might be argued that this is neither good nor bad, and that if candidates wish to risk repeated test failures, they have the right to do so. However, in view of the benefits of increasing the probability of passing the test (see (c) above), it is useful to consider how the current pass rate may have arisen and what might be done to increase it.

An obvious possible explanation is that the rather low pass rate results from candidates attempting to minimise the total cost of obtaining a driving licence by coming for test too early in the hope that they will pass. No doubt this is one component. The report also indicates that instructors’ optimism about the chances of a candidate passing may be another, with expectations of parents, and competition with friends, also playing a part. However, the evidence presented in Section 7 indicates that there are other factors at work that might be addressed by changes to the test itself. These include candidates’ beliefs that their moderate probability of passing arises not so much from limitations of their own driving, but from factors outside their control. The generality of this finding, the reasons for it, and possible solutions, need further investigation. If it were possible to increase candidates’ perceived control over test outcome, they might then choose to delay coming for test until they had a higher probability of passing, especially if the consequences of failure were increased by (for example) increasing the cost of retests or the compulsory minimum delay between test and retest. However, research is needed to explore the effect of increased lessons and practice on test performance and rates of making serious and dangerous faults. This would extend previous DfT research on the relationship between error rates, training and practice.

Again, other interventions designed to increase the amount of training and practice accumulated by learner drivers may be more effective at increasing the pass rate than changes associated with the test itself.

11.9 Feedback

The driving test is an opportunity for candidates, instructors and supervisory drivers to receive feedback from the examiner. This could cover the candidates’ competence and limitations, and the importance of motivational/attitudinal influences on driving. Only a short feedback session at the end of the test is included at present and this only for candidates who fail the test. Also, despite encouragement from DSA, few instructors choose to be present in the car during the test, or to listen to the examiner’s feedback to the candidate after the test. The reasons for the apparent reluctance merit further examination. Instructors appear, from the evidence presented in the report, to be optimistic about candidates’ probabilities of passing. This, and the fact that it is desirable for candidates (especially those who pass the test) to appreciate their own limitations, means there is a good argument for increasing the emphasis on feedback.
Candidates may not be receptive to detailed feedback immediately after the test, and the provision of instructors’ time, examiners’ time and debriefing facilities would be expensive to organise. Nevertheless, a lot of emphasis is given to test feedback in some other countries, and it would seem worth exploring further in Britain. It may be appropriate to review the subject of test feedback anyway if other changes – particularly involving certified practice and the widespread or mandatory keeping of a learner driver logbook – are made to the licensing system.

12 Overall conclusions and recommendations

1 Tightening the failure criterion for ‘driving faults’
There is a valid argument for tightening the failure criterion for ‘driving faults’ from its current limit of 16. Such a change would cause a group whose driving is intrinsically less safe than that of others\(^\text{23}\) to fail the test, and therefore to have more training and practice before driving unsupervised. However, this argument is finely balanced: in practice this group of people tend to do less driving than ‘lower-faults’ drivers, at least in the first six months of post-test driving. So, although their driving is intrinsically less safe, their actual number of accidents is similar to that of lower faults drivers.

3 Persuading people to delay coming for test
It is desirable to find ways of persuading people to accumulate more driving experience before coming for test. This would reduce post-test accident liability, and should also increase the test pass-rate. However, increasing test fees and delaying re-tests might not be effective in persuading L-drivers to delay coming for test in order to improve their probability of passing. More needs to be known about the ability of candidates to reduce their error-rates (and increase their test pass-rates) by taking additional training or practice. Learner drivers’ perceptions their ability to influence test outcome in this way also need to be better understood. The apparent optimism of both candidates and instructors regarding the probability of passing the driving test also merits further examination.

3 The content of the practical driving test
The possibility of broadening the test to include independent driving skills such as choosing where and how to conduct manoeuvres, dealing with missed turnings, finding a route using direction signs, and route planning, merits consideration. The potential for strengthening the assessment of hazard perception during the practical driving test, in order to complement the computerised hazard perception test that is now included in the theory test, is worth exploring further. The growing realisation of the importance of drivers’ attitudes and goals as influences on behaviour (Hatakka et al., 2002), and the need for drivers to be aware of these influences and how to recognise and control them has implications for both driver training and testing. It may not be feasible to incorporate such aspects into the pass-fail criteria for the test but, they could, in principle, be incorporated in the practical test to emphasise their importance (Baughan et al. 2005).

4 The test marking system
Tests of the types used in Sweden, the Netherlands and Western Australia place less reliance on the making and recording of individual errors than the British test, and more reliance on examiners’ judgements of competence. If further improvements to driver testing are sought, it would be useful to investigate these tests alongside the British test to assess whether they offer benefits, and whether these are likely to be sufficiently great to justify the investment that would be needed if elements of such tests were to be introduced in Britain.

5 The test as a learning opportunity
The potential for making more use of the driving test as a learning opportunity should be further assessed by studying practice in countries that already give more emphasis to test feedback. The benefits and feasibility of encouraging more instructors to be present in the car during the test, or to listen to the examiner’s feedback after the test, merit further examination.

6 Consistency of test standards
Between-centre differences in pass-rates merit further examination to discover whether they represent undesirable differences in route/traffic demands or assessment standards, or whether they are explained by differences in the competence of candidates coming forward for test – perhaps associated with variations in access to training and supervised practice. Information on the amounts of training and practice accumulated by candidates at different centres or regions would be useful here.

A study of the extent to which drivers restrict their early driving to the area where they take their test would be instructive in helping to assess the importance of regional differences in test conditions.

If there are identifiable types of test centre at which candidates come for test particularly poorly prepared, then low pass-rates at such centres would imply that the test is doing a useful job in screening out such candidates. However, policies to improve candidates’ preparation at these centres may be needed.

The effect of the choice of reversing manoeuvres on pass rate is complex, involving interactions between difficulty of the manoeuvre, time of day of test, and possibly other variables. Nevertheless, the rather large differences between the manoeuvres in terms of the probability of making a serious or dangerous fault mean that this area merits further examination.

\(^{23}\) *i.e. their accident liability (statistically expected number of accidents) in the first six months of driving would be higher than that of lower-faults drivers once the effects of mileage and age had been adjusted for.*
Other aspects of the training/testing/licensing system

Changing the driving test should not be seen as the only, or necessarily the best, way of improving novice driver safety. The licensing system itself can be modified to encourage or require learner drivers to take more and better training and practice before driving solo, and to influence their behaviour and exposure to risk during the early months of solo driving. Such changes have been widely and successfully applied in other countries and have been considered in Britain in the Department for Transport’s consultation document ‘Introducing a More Structured Approach to Learning to Drive’ (DTLR, 2002).

Acknowledgements

The work described in this report was sponsored by Road Safety Division of the Department for Transport and carried out in Safety Group, TRL.

The authors would like to thank:

Elaine Forsyth, the Department for Transport’s Research Programme Manager for this project.
Robin Cummins, DSA Chief Driving Examiner during the life of this project; Trevor Wedge, then DSA Deputy Chief Driving Examiner and Manager for External Training Development and now Chief Driving Examiner; and the many DSA examiners and other staff who gave invaluable help with the study.
Diane Firth and Linda Goffey of Arena Research and Planning for recruiting respondents and conducting the exploratory interviews with test candidates reported in Section 7.
John Hanson and Kate Riches of the National Foundation for Educational Research (NFER) for administering the fieldwork for the survey described in Section 7.
Phillip Crofts of PAS Ltd for conducting the group discussions of examiners and instructors summarised in Section 4.2.
The many test candidates, novice drivers and driving instructors who participated in the studies.
Ex-TRL colleagues Christian Palmer and Lee Smith who participated in aspects of the study.

References


Sagberg F (2001a). *Novice drivers’ crash risk before and after the age limit for driver training in Norway was lowered from 17 to 16 years. TOI report 498/2001. Oslo: Institute of Transport Economics.*

Sagberg F (2001b). Personal communication.


The influence of the driving test on training limits both the predictive validity of the driving test and our ability to measure it. Only a fully implemented test will induce changes to L-drivers’ training and practice, and these changes will take time. In the process of inducing them, a test may reduce or remove its own predictive validity by reducing the range of driving performance (and accident liability) of candidates. Cognitive changes may then be achieved in the course of learning or by post-training practice, and these changes will take time. Because of methodological constraints, it is often not possible to determine the effect of a test designed to influence novice drivers on their behaviour post-test.

Consequences of unreliability

The implications of poor test–retest reliability are not straightforward. To examine them it is useful to consider an extreme case in which a pass-fail driving test has zero test–retest reliability for current candidates. A test with zero test–retest reliability for actual candidates (ie 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, it 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 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 test–retest 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.

If candidates do tend to choose to come for test with only a moderate subjective probability of passing it will be difficult to achieve high test–retest reliability by improving the test alone. Candidates may well adapt to the new test such that they continue to come with a moderate subjective probability of passing. It may be necessary to manipulate the probability of passing that the candidate chooses to reach before coming forward for test. Section 7 explores whether this is likely to be achievable through measures that increase the penalty for failure – eg By increasing test fees or the delays between test and retest.

---

1 i.e. for candidates who choose to come forward for test under the prevailing conditions.
Appendix B: Rates and test reliability

B.1 A simplified model of a driving test

Some insight into the effect of individual candidates’ underlying competence on the test–retest reliability of a driving test, and on its ability to identify and fail people with unacceptably poor underlying levels of competence, can be gained by examining a simplified model of a driving test.

In this simplified model, the level of driving competence reached by a test candidate is characterised by that individual’s underlying error rate – ie the average number of errors per time period that would be made during a large amount of driving. The test aims to fail candidates whose underlying error rate exceeds some threshold. To do this it observes the candidate driving for a given period, counts the number of errors actually made, and fails candidates who reach or exceed a stated number of errors. For the purposes of the model, a candidate’s errors are assumed to be generated by a Poisson process – ie a candidate’s underlying error rate is the mean of a Poisson distribution. Such a model can be used to explore test–retest reliability (measured as proportion of test–retest agreements) and probability of misclassification (ie failing candidates who have expected error rates below the desired threshold, or passing those with error rates above the threshold). It can show how these are likely to be influenced by test duration, underlying error rate, pass mark, and threshold error rate.

Table B.1 shows cumulative Poisson probabilities (the probability that r or more events will occur in a stated interval of time). Suppose that a test is intended to fail candidates with expected error rates per test of 1 or more, and that the failure criterion is one or more actual errors committed during the test. The first row of Table B.1 shows the test will correctly fail at least 98% of candidates with expected error rates of 0.1 or less (the table entry shows that 9.5% of candidates with an expected error rate of 0.1 will fail). However, a rather high proportion (39%) of candidates with an expected error rate of 0.5 (ie half the target threshold rate) will be incorrectly failed, and (100–86.5) = 13.5% of candidates with an expected error rate of 2 (double the target threshold) will be incorrectly passed.

Maximum test–retest agreements (ie assuming no other sources of unreliability) for candidates with given expected error rates can also be calculated from Table B.1.

For example, candidates with an expected error rate of 0.5 per test have a probability of 0.394 of failing each test. Their probability of obtaining agreeing results in a test–retest study would therefore be the probability of getting two passes + the probability of getting two fails, ie (1-0.394) 2 + 0.394 2 = 0.522. Even with no other sources of unreliability acting, the test would yield only 52% test–retest agreements for such candidates. The first row of Table B.2 shows the results of similar calculations for other expected error rates, and indicates that test–retest agreements remain rather low over a wide range of expected error rates.

Table B2 Percentage of test – retest agreements for different expected error rates

<table>
<thead>
<tr>
<th>Expected errors per test (Poisson intensities)</th>
<th>0.1</th>
<th>0.2</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 error</td>
<td>83%</td>
<td>73%</td>
<td>52%</td>
<td>53%</td>
<td>77%</td>
<td>90%</td>
<td>96%</td>
<td>100%</td>
</tr>
<tr>
<td>2 errors</td>
<td>99%</td>
<td>96%</td>
<td>84%</td>
<td>61%</td>
<td>52%</td>
<td>68%</td>
<td>83%</td>
<td>99%</td>
</tr>
</tbody>
</table>

Given a distribution of expected error rates for the population of candidates, the model can be used to predict the proportion of agreements in a test–retest study, and the proportions of candidates who will be correctly failed and incorrectly failed by the test.

Tables B.3 and B.4 show the results of doing this for a distribution fairly similar to the one actually found in a large study of test errors carried out as part of the project. The failure criterion is 1 or more errors and, for Table B.4, the hypothetical test here is aiming to fail candidates with an expected error rate per test period of one or more.

Table B3 Simple model: test–retest reliability

<table>
<thead>
<tr>
<th>First test</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>Fail</td>
<td>16</td>
<td>50</td>
</tr>
<tr>
<td>Pass</td>
<td>34</td>
<td>66</td>
</tr>
</tbody>
</table>

Table B4 Simple model: correct and incorrect outcomes

<table>
<thead>
<tr>
<th>Candidate should</th>
<th>Test outcome</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>25</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>9</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>34</td>
<td>66</td>
<td></td>
</tr>
</tbody>
</table>

Cell entry = % of total candidates.

Table B1 Poisson probabilities

<table>
<thead>
<tr>
<th>Poisson intensity (expected events per time interval)</th>
<th>r</th>
<th>0.1</th>
<th>0.2</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.095</td>
<td>.181</td>
<td>.394</td>
<td>.632</td>
<td>.865</td>
<td>.950</td>
<td>.982</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.005</td>
<td>.018</td>
<td>.090</td>
<td>.264</td>
<td>.594</td>
<td>.801</td>
<td>.908</td>
<td>.997</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.001</td>
<td>.014</td>
<td>.080</td>
<td>.323</td>
<td>.577</td>
<td>.762</td>
<td>.986</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.002</td>
<td>.019</td>
<td>.143</td>
<td>.353</td>
<td>.567</td>
<td>.958</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cell entry = probability that r or more events will occur in the specified time interval.

2 If the only source of unreliability is their non-zero error rate.
The pattern of test–retest agreements is not dissimilar to that shown in Table 8 for the real driving test. Given the assumptions being made, this degree of similarity is probably spurious. However, the model result does show that an ‘error counting’ test based on observing infrequent errors is likely to have high levels of test–retest unreliability because it is poor at estimating underlying error rates.

Table B.4 shows that the model test makes an ‘incorrect’ decision for \((17+9) = 26\%\) of candidates. Like the test–retest agreements, this result depends on the choice of the population distribution of expected error rates – ie the distribution of test candidates’ driving ability.

There are, of course, degrees of incorrectness in pass-fail decisions. It does not really matter if people with expected error rates close to the test threshold are misclassified by the test, but it is important for the test to fail people with much higher error rates and to pass people with much lower ones. As already discussed, Table B1 gives some insights into the ability of a test to do this.

### B.2 Improving the reliability of the driving test

As discussed in Section 2.3, the driving test could still be performing its main function of inducing training and practice despite the pattern of test–retest disagreement shown in Table 2. Nevertheless, good reliability is desirable for the reasons outlined. Also, a test with good reliability and good discriminatory power is likely to be more effective in inducing training and practice.

‘Wrong decisions’ in the simplified model of the test occur because of the difficulty inherent in using observed errors to estimate underlying error rates. Test–retest disagreements stem from a similar problem – the fact that a candidate with a constant underlying error rate will produce different numbers of errors on different occasions.

In principle, these problems could be addressed by increasing the numbers of errors observed during the test to increase the precision with which it is able to estimate the underlying error rate. This could be done by increasing the duration of the test or by basing the test on errors that are less severe and more frequent.

The likely effects of increasing the duration of the test can be explored using the model, on the simplifying assumption that a candidate’s expected error rate per unit time will remain unchanged. Under this assumption, if the test duration is doubled then a candidate’s expected number of errors per test will also double. Table B1 shows the effect that this would have on candidates at each expected error rate. For example, a candidate with an expected error rate per test of 2 in the table has a probability of failure (ie of making one or more errors) of 0.865. In a doubled test, the same candidate would have an expected error rate per test of 4. Assuming that the failure criterion for the double length test is also doubled\(^3\) (to two or more errors), row 2 of the table shows that the candidate would now have a probability of failure of 0.908.

Similarly, for a candidate with an expected error rate of 0.1, the probability of an (incorrect) failure would reduce from 0.095 to 0.018. The test would become somewhat better at failing people with high error rates, and passing people with low ones.

The second row of Table B.2 shows the effect on test–retest reliability. Candidates with expected error rates of 2 and 0.1 per standard length test would have rates of 4 and 0.2 in a doubled test. Their test–retest agreements would improve from 77\% to 83\%, and from 83\% to 96\% respectively.

One disadvantage of the above approach is the cost of a longer test. The failure criterion for this type of test can only be an integer number of errors and, if it were to change to two errors, duration would have to be approximately doubled to maintain the current overall pass rate and test standard. A second problem is that the single error which is currently sufficient to produce a test failure is labelled as a serious or dangerous fault: one that involves potential or actual danger. A longer test that failed someone for making two such faults would be more effective than the current test at identifying candidates with a high underlying rate of serious or dangerous faults. However, it would involve passing candidates who make a single serious or dangerous fault. DfT and DSA have made it clear that it would be unacceptable to pass candidates who make such a fault during the test.

Section 6.5 discusses alternative approaches to increasing reliability.

---

\(^3\) If it the failure criterion remained unaltered, the test would become much more severe, and the pass rate would drop sharply.
Abstract

Novice drivers have a much higher accident liability than more experienced drivers. One approach to improving novice driver safety is to modify the driving test to induce learner drivers to accumulate more or better training and experience, and to screen-out drivers who have not yet reached a standard acceptable for solo driving. The Department for Transport (DfT) commissioned TRL to assess the scope for such improvements. A number of potential improvements were identified, and some were implemented in May 1999. To develop an understanding of how the test might be further improved, studies were made of the consistency of candidates’ performance at the time they come for test, the way in which candidates decide when to come for test, the literature on novice driver accidents and its implications for driver testing, the influence of test route and other test characteristics on pass rates, and the relation between numbers and types of faults made during the test and subsequent accidents.

In Britain, the driving test is currently the main tool for inducing learner drivers to build up their competence by training and experience before driving solo. However, to achieve the desired improvements in safety it may be necessary to make other changes to the training/testing/licensing system so that it becomes less reliant on the driving test itself. The project included a review of licensing systems in other countries with a view to identifying measures that might be effective in Britain.

Related publications

TRL558  The development of hazard perception testing by G B Grayson and B Sexton. 2002 (price £35, code AX)
TRL542  In-depth accident causation study of young drivers by D D Clarke, P Ward and W Truman. 2002 (price £35, code AX)
TRL529  Graduated driver licensing - a review of some current systems by C J Baughan and H Simpson. 2002 (price £30, code A)
TRL527  Novice driver accidents and the driving test by G Maycock. 2002 (price £35, code AX)
TRL525  Monitoring and evaluation of safety measures for new drivers by H Simpson, L Chinn, J Stone, M Elliott and J Knowles. 2002 (price £30, code A)
TRL326  Drivers’ speed choice: an in-depth study by A Quimby, G Maycock, C Palmer and G B Grayson. 1999 (price £35, code E)
TRL275  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 by G Maycock and E Forsyth. 1997 (price £45, code H)
RR338  Cohort study of learner and novice drivers, Part 1: Learning to drive and performance in the driving test by E Forsyth. 1992 (price £45, code H)
RR315  The accident liability of car drivers by G Maycock, C R Lockwood and J Lester. 1991 (price £30, code C)
CT77.2  Driver training update (2000-2001). Current Topics in transport: selected abstracts from TRL Library’s database (price £30)

Prices current at December 2005

For further details of these and all other TRL publications, telephone Publication Sales on 01344 770783, email: publications@trl.co.uk, or visit TRL on the Internet at www.trl.co.uk.