Published Project Report
PPR312

Creating the future of transport

# Situational judgement in driver training and assessment: a literature review

S Helman

**Transport Research Laboratory**

# SITUATIONAL JUDGEMENT IN DRIVER TRAINING AND ASSESSMENT: A LITERATURE REVIEW

**by S Helman (TRL Limited)**

**Prepared for: Project Record: PPRO 04/045/004 LOT3 / 018**

**LITERATURE REVIEW: SITUATIONAL JUDGEMENT ELEMENT IN THE DRIVER TRAINING AND ON ROAD ASSESSMENT PROTOCOL: BUILDING AN EVIDENCE BASE**

**Client: Driving Standards Agency (DSA) Department of Applied Research (Bob Hannigan)**



This Published Report has been prepared for Driving Standards Agency (DSA) Department of Applied Research.

The views expressed are those of the author and not necessarily those of Driving Standards Agency (DSA) Department of Applied Research.

| | Approvals |
|---|---|
| **Project Manager** | **Su Buttress** |
| **Quality Reviewed** | **Britta Lang** |

When purchased in hard copy, this publication is printed on paper that is FSC (Forest Stewardship Council) registered and TCF (Totally Chlorine Free) registered.

# CONTENTS

# Executive summary

The DSA's vision ("Safe Driving For Life") forms the basis of the DSA's mission to contribute to the Government's 2010 Road Casualty Reduction Targets set out in 2000. The DSA envisages its major contributions to this road safety initiative to be in the areas of:

- Setting standards for drivers, riders and instructors;
- Driver education and the provision of learning resources;
- Registering and supervising quality assured instructors;
- Modern, effective and efficient assessments conducted as computer-based and practical tests.

In this report, work related to modern, effective and efficient assessments is discussed; and in particular, possible changes to practical driving training and assessment.

The assessment of a candidate on their practical driving test in Great Britain has traditionally been based on what the examiner observes during the test. The measurement of errors (driving faults and serious/dangerous faults) serves as an objective means of passing or failing the candidate. Recently a number of new approaches to driver assessment have been suggested. These range from radical changes to the current system such as moving to graduated licencing, to smaller changes such as the introduction of verbally-based 'situational judgement' assessments and commentary driving during the driving test, or the introduction of independent driving elements.

DSA tasked TRL with a review of the evidence for two specific techniques that may merit inclusion in future versions of the practical driving test in Great Britain: situational judgement testing, and commentary driving. Both techniques have been suggested as ways of assessing the higher cognitive processes that underlie observed driving behaviours during practical driver training and assessment.

Situational judgement testing involves asking a candidate questions about a particular manoeuvre that he/she has just carried out, is about to carry out, or about hypothetical (but relevant) situations. It is possible that such testing would allow the examiner to understand the candidate's mental decision making process (leading up to, and during the actual manoeuvre). It may also allow the driving instructor to assess whether the candidate has missed something, acted deliberately, etc. Situational judgement testing may thus help uncover the reasons why a driver is behaving as observed, and therefore help generalise from the 'maximal' behaviour observed in the test to a driver's likely 'typical' behaviour during later unsupervised driving. Similarly, the inclusion of situational judgement questioning during learner driver training may help driving instructors tailor their tuition to individual drivers' needs as they would better understand their learners' perceptions and decision making processes and the way these may be affected by social, motivational or knowledge-related factors.

An alternative way of accessing the candidate's thought processes would be to have them engage in commentary driving. Commentary driving comprises the verbal report of road hazards and situations as driving is taking place.

Both techniques have been included in driver training and assessment regimes in other countries, and are widely covered in related academic literatures. Therefore theoretical (i.e. academic) and applied literatures were reviewed, including literatures on: employment selection; situational awareness; hazard perception; general cognitive psychology; and evidence from implementation of the two techniques in other countries. Additionally, interviews were held with police driving instructors regarding the practical implementation of commentary driving.

The following research question formed the focus of the review:

> "What evidence is there relevant to the suitability of situational judgement testing and the technique of commentary driving for inclusion in future practical driving training and assessment?"

## Definitions

Four key concepts in driving are defined in the report, in order to help avoid confusion in any future work developing the techniques of situational judgement testing and commentary driving: situational judgement; situational awareness; hazard perception; and commentary driving.

Situational judgement is a concept that has its origins in the employment selection literature. Put simply, measures of situational judgement present job candidates with job-relevant situations and ask them to demonstrate appropriate knowledge or to select appropriate behaviours in those situations. Situational judgement in driving is defined as:

> "the application of driving-related knowledge and behavioural tendencies (e.g. personality, attitudes, beliefs etc.) to the assessment of traffic situations, including knowledge of the appropriate driving behaviours in those situations."

The related concepts of situational awareness and hazard perception are defined respectively as:

> "the perception of other road users and features of the road environment, the comprehension of their meaning in isolation and/or in combination, and the prediction of their status in the near future;" and
>
> "the ability to identify potentially dangerous traffic situations as early as possible."

Having defined situational awareness and hazard perception in this way the review is able to focus on what is different about situational judgement—namely that it goes beyond the skill of detecting hazards, to include the knowledge of appropriate behaviours and attitudes to the driving task.

Commentary driving is defined as:

> "the verbalisation of at least some of the driving-related contents of awareness, while actually driving through a situation."

It is also acknowledged that commentary can be delayed (i.e. done after driving through a situation).

## Findings

The key findings in the review are as follows:

- Situational judgement is usually measured through the administration of formal Situational Judgement Tests—usually in the format of a description of a situation, with alternative behaviours (varying in appropriateness as rated by experts) for respondents to choose from.
- Such tests require considerable effort to create and administer correctly, in order that the data obtained are valid predictors of performance in the domain in question.

- Situational judgement testing has been shown to have several advantages over traditional selection instruments such as interviews, and cognitive ability testing, including better prediction of job performance.
- This increased predictive power is probably due to the fact that Situational Judgement Tests assess the behavioural tendencies and personality factors needed to succeed in a job, as well as the knowledge required.
- Situational judgement tests appear to be culturally fairer to minority (e.g. ethic) groups, than tests of cognitive ability.
- The way in which situational judgement test items are worded is especially important since it can alter what the test is measuring (knowledge, or behavioural tendencies).
- Formal situational judgement tests are unlikely to be suitable for use in practical driver training and assessment. They are more likely to have use in theory testing, where control can be kept over the situations presented.
- There are some examples of situational judgement-like tests being introduced in the practical driver assessment of other countries; however there is, as yet, no validated test of situational judgement for driving.
- Even if situational judgement testing is not currently feasible as a pass/fail metric, there is still value in inclusion of such testing in driver training and assessment, since it draws attention to the attitudes and behavioural tendencies that new drivers should be showing in order to be safe on the road.
- By drawing the attention of drivers to their attitudes and perceived skills, those who are badly calibrated (i.e. by overestimating their skills and underestimating the demands of the driving task) should improve and road safety benefits should follow.
- The concepts of hazard perception and situational awareness are related to situational judgement, in that they also address drivers' mental processing and awareness of road hazards. However, we argue that situational judgement can be thought of as distinct from situational awareness and hazard perception, since situational judgement goes beyond awareness of hazardous road situations and specifically includes attitudes and behavioural tendencies regarding driving.
- It is also argued that situational awareness, and hazard perception in driving, can be seen as essentially equivalent; i.e. hazard perception can be seen as situational awareness for driving situations.
- Commentary driving, when carried out concurrently with the driving task itself, is difficult, and associated with a decrease in driving performance. This is true even for highly skilled (police) drivers.
- Another challenge for commentary driving as a technique is that driving test candidates may not be able to perform it sufficiently well for it to be of use; candidates may prioritise the driving task at the expense of the commentary. Furthermore they may not be able to put some of their driving knowledge into words—the research suggests that people are bad at verbalising knowledge that underlies complex skills such as driving.

## Recommendations

The following recommendations are made based on the main findings from the review:

- Situational judgement testing has potential for use in driver training and assessment. If worded appropriately, such testing will be especially suited to probing the behavioural tendencies of drivers in given situations as well as their

knowledge of road hazards (i.e. hazard perception or situational awareness). Therefore, DSA should consider developing and evaluating situational judgement tests for driving.

- If this is to be done for practical driver assessment, then the practical implications of such a technique will need to be assessed before implementation. This includes assessing the degree to which such tests (and their scoring) can be standardised within a dynamic and changing practical test environment, and the degree to which time constraints make the tests suitable for use in the time available on the practical test.
- Less formal methods of testing situational judgement, such as verbal descriptions of traffic situations experienced during training and on test ('delayed commentary') are likely to be the most suitable for the practical driving test, while formal situational judgement tests are mainly suitable for theory testing where there is an opportunity to completely control contents.
- If it is the case that less formal methods used on practical driving tests cannot be standardised, then they should not be used for 'assessment' and 'pass/fail' decisions; rather they should be used to draw attention to the issue of 'calibration', so that learner drivers can be encouraged to match their driving behaviour to the driving-task demands, and the risks present in the driving environment.
- The introduction of some kind of situational judgement testing (formal or informal, assessed or not-assessed) should be accompanied by consideration of the good-practice methods for test creation outlined in the review—especially item wording. Ignoring such issues will undermine the effectiveness and validity of such testing in predicting safe driving outcomes.
- There is a wide range of evidence that verbal secondary tasks (especially those that involve production of speech) adversely impact on driving performance. Commentary driving (verbal commentary while actually driving a vehicle) is therefore likely to have deleterious effects on performance, including effects on car control and hazard perception, at least during early stages of learning, when car-control skills are still being acquired. Feedback from Police driving instructors echoed this point of view. Any development of commentary driving should therefore proceed under the premise that it should either be carried out after a particular situation, rather than during actual manoeuvring, or it should be carried out by candidates while someone else (i.e. driving instructor) drives the car.
- Further work on assessing the likely benefits of such 'delayed' commentary driving (even if only as a method of measuring situational judgement or hazard perception 'on the road') is necessary, along with work to quantify the level of interference with the driving task that results from commentary driving. This work will permit decisions to be made as to the suitability and usefulness of commentary driving in the future.

# Abstract

Situational judgement testing and commentary driving are two methods that can be used to assess the higher cognitive processes that underlie observed driving behaviours. The research literature is reviewed, and advice sought from expert police drivers, to assess the suitability of these two methods for use in practical driver training and testing in Great Britain (GB). First the terms 'situational judgement' and 'commentary driving' are defined within the driving context. Situational judgement is defined as "the application of driving-related knowledge and behavioural tendencies (e.g. personality, attitudes, beliefs etc.) to the assessment of traffic situations, including knowledge of the appropriate driving behaviours in those situations", and is distinguished from the related terms of 'hazard perception' and 'situation awareness'. Commentary driving is defined as "the verbalisation of at least some of the driving-related contents of awareness, while actually driving through a situation". Key findings from the relevant literatures—related to the practical use of these methods in the GB driver training and testing protocol—are then listed. It is concluded that neither method is suitable as a 'pass-fail' instrument in the practical driving test in GB, although situational judgement testing has considerable promise as a method for use in the driving theory test, as long as good-practice procedures for test creation are followed. The focus of situational judgement testing should be on the assessment of safe 'behavioural tendencies' in driving as well as 'knowledge' of what constitutes safe and appropriate behaviour. It is suggested that commentary driving would not be suitable for learner drivers while they are in control of the vehicle, due to possible deleterious effects on driving performance.

# 1 Background and project aims

1.1 The core aim of the Driving Standards Agency (DSA) is to help deliver improvements in road safety by influencing driver behaviour through setting the standard for safe drivers and trainers, educating drivers, supervising trainers and assessing the standard of driving at test. The DSA seeks to work in close partnership with public and private stakeholders in order to develop its role in delivering higher driving standards in Great Britain.

1.2 These goals are reflected in the DSA's vision "Safe Driving For Life" and its mission is to contribute to the Government's 2010 Road Casualty Reduction Targets set out in 2000. The DSA envisages its major contributions to this road safety initiative to be in the areas of:

- Setting standards for drivers, riders and instructors.
- Driver education and the provision of learning resources.
- Registering and supervising quality assured instructors.
- Modern, effective and efficient assessments conducted as computer-based and practical tests.

1.3 In this report, we discuss work related to the fourth of these; specifically to possible changes to practical driving training and assessment.

1.4 TRL has been tasked with reviewing evidence for two techniques that may merit inclusion in future versions of the driving test in Great Britain: situational judgement testing, and commentary driving. These two techniques have been included in driver training and assessment regimes in some other countries, and also have sizeable coverage in the related academic literature. Both techniques have been suggested by DSA in the research specification for this project as ways of assessing the higher cognitive processes that underlie observed driving behaviours during practical driver training and assessment.

1.5 The assessment of a practical test candidate has traditionally been based on what the examiner observes. Asking the candidate questions about a particular manoeuvre that he/she has just carried out would allow the examiner to understand the mental decision making process of the candidate (leading up to, and during the actual manoeuvre), and whether the candidate has missed something, acted deliberately, etc. (CIECA, 2007). Baughan and Keskinen (2006) have pointed out that the inclusion of situational judgement assessment can help uncover the reasons why a driver is behaving as observed, and so help generalise from the 'maximal' behaviour being observed in the test to a driver's likely 'typical' behaviour during later unsupervised driving. Similarly, the inclusion of situational judgement questioning during learner driver training could help driving instructors to tailor their tuition to individual drivers' needs as they would better understand their learners' perceptions and decision making processes and the way these may be affected by social, motivational or knowledge-related factors.

1.6 The aim of this review is therefore to examine and assess evidence related to situational judgement testing and the technique of commentary driving, to deduce their suitability for inclusion in future practical driving training and assessment. Evidence from both theoretical and applied literatures is reviewed.

1.7 The remainder of this section provides the background to the review and its aims, and also sets out the structure of the remainder of the report.

## The problems faced by novice drivers

1.8 Research has revealed that one of the best predictors of accident risk for drivers is a lack of driving experience (e.g. Maycock, Lockwood and Lester, 1991; Maycock, 2002). Accident risk drops sharply during the first six months of post-test driving, and then continues to drop for the next two to three years. Put simply, as soon as new drivers embark upon their post-test driving, they encounter the highest-risk period of their driving careers.

1.9 There is wide consensus that the heightened accident risk of new drivers is due largely to a lack of higher order cognitive skills such as hazard perception, and the fact that they hold beliefs and attitudes that are linked to risky behaviours on the road (e.g. Engstroem et al., 2003; Hatakka et al., 2002; Quimby, Maycock, Carter, Dixon and Wall, 1986; Evans, 1991; Hull and Christie, 1993; McKenna and Horswill, 1999). For example, hazard perception skill is related to experience (e.g. McKenna and Crick, 1991; McKenna and Horswill, 1999) and has been shown to be related to accident risk (e.g. Quimby et al., 1986; Hull and Christie, 1993; McKenna and Horswill, 1999). Unsafe attitudes and beliefs of drivers have also been shown to be related to accidents (see Engstroem et al, 2003, for a review). Horswill and McKenna (2004) among others point out that this link with accident risk overall does not appear to be present for lower-order perceptual skills (e.g. Quimby et al., 1986) or vehicle control skills (see Evans, 1991 for a review).

1.10 There is currently no formal process in GB driving assessment for probing directly the cognitive processes underlying hazard perception during practical training and assessment. If such knowledge were being probed explicitly, then driving instructors and examiners may be able to make use of it during practical training and assessment. Additionally, there is currently no formal method for assessing higher levels (goals for life and skills for living; and goals and context of driving) of the Goals for Driver Education (GDE) matrix (Hatakka, Keskinen, Gregersen, Glad and Hernetkoski, 2002; Siegrist (ed.), 1999—see Appendix A), which is used widely as a standard framework for explaining the multiple hierarchical levels of skills, behaviours, attitudes and beliefs that contribute to driver behaviour (and thus safety) on the road.

1.11 Situational judgement testing and commentary driving are two techniques that may permit assessment of such higher-order processes.

## Why situational judgement testing?

1.12 Situational judgement tests (SJTs) and interviews are used extensively in employment selection and assessment. They present scenarios from a specific (i.e. job related) context, and ask respondents to pick a response to the scenario from a set of alternatives. An example of an SJT item is given below (as cited in McDaniel and Whetzel, 2005a, p516):

*A man on a very urgent mission during battle finds he must cross a stream about 40 ft wide. A blizzard has been blowing and the stream has frozen over. However, because of the snow, he does not know how thick the ice is. He sees two planks about 10 ft long near the point where he wishes to cross. He also knows where there is a bridge about 2 miles downstream. Under the circumstances, he should:*

*A. Walk to the bridge and cross it.*

*B. Run rapidly across the ice.*

*C. Break a hole in the ice near the edge of the stream and see how deep the stream is.*

*D. Cross with the aid of the planks, pushing one ahead of the other and walking on them.*

*E. Creep slowly across the ice.*

1.13 Situational-based measures such as SJTs have been shown to have several advantages over more traditional assessment techniques, including:

- they show less adverse impact against minorities compared to other psychometric tests such as cognitive ability tests;
- they use measures that directly assess relevant behaviours;
- they can be administered in bulk, either via pen and paper or on-line; and
- they are more acceptable and engaging to candidates compared to cognitive ability tests since scenarios are based on real incidents.

1.14 Given the apparent promise of situational based measures in employment assessment, it is worth considering whether they can be included in driver training and assessment, to compliment other methods currently used in GB driving training and assessment.

## Why commentary driving?

Commentary driving (also known as concurrent verbalisation or the 'think aloud technique') refers to the method of having a driver give a running commentary of his or her thought processes while driving. It is used extensively in advanced police driver training, to enable drivers to learn to consider as many of the possible hazards in a driving situation as they can, and to demonstrate their understanding of a driving situation to instructors. Clearly, commentary driving (assuming it can be done safely and effectively) is another technique that may enable the real-time measurement of learner drivers' thought processes, so that instructors and examiners may make inferences from these thought processes to the behaviours exhibited while driving, and the learner drivers' understanding of driving situations. If commentary cannot be done safely concurrently with driving, then it is possible that commentary before or after hazardous road situations may be more practical.

## Structure of the report

1.15 The remainder of this report is split into four main sections.

- Section 2 offers precise definitions of situational judgement, and commentary driving, in the context being discussed. The related concepts of situational awareness and hazard perception are also defined, so that the ways in which they differ from situational judgement can be made clear. These definitions set a clear boundary for the review.
- Section 3 presents evidence for the effectiveness of situational judgement testing, by studying the employment selection literature. Then evidence for the suitability (or not) of situational judgement testing for driving training and assessment is reviewed, including the examination of other implementations of such testing in the driver training and assessment of other countries. A large part of Section 3 is also focused on the nature of formal situational judgement tests—the usual method by which situational judgement is measured.

- Section 4 focuses specifically on commentary driving. It examines the theoretical and evidential basis for its suitability for use with learner drivers as a means of gaining insight into their thought processes, and also examines evidence for the effectiveness of commentary driving as a training tool and as an indicator of the contents of awareness in expert police drivers.
- Section 5 draws overall conclusions based on the evidence discussed in previous sections, and gives recommendations to assist DSA in making informed, evidence-based decisions regarding the inclusion (or not) of situational judgement testing and/or commentary driving in future GB driver training and assessment.

1.16 At the beginning of each of Sections 2, 3 and 4, a callout box of bullet points is included, to highlight the key findings and recommendations of the section. The overall conclusions and recommendations in Section 5 are also presented in this format.

1.17 Throughout the report, the focus is kept on the two main issues of situational judgement testing and commentary driving. However, additional material is reviewed from literatures on hazard perception, situational awareness, and general cognitive psychology, where relevant. The review is not, however, designed to be an in-depth review of these additional fields. The interested reader is referred to Banbury and Tremblay (2004), for a more detailed review of situational awareness, and to Horswill and McKenna (2004) for a more detailed review of hazard perception.

1.18 Details on the literature databases and methods used for the review can be found in Appendix B.

# 2 Definition of concepts


**Summary of Section 2**

- The concepts of situational judgement, commentary driving, situational awareness, and hazard perception are defined.
- Situational awareness and hazard perception can be seen as equivalent in driving—they are both ways of describing how drivers are aware of elements (e.g. vehicles, pedestrians, road rules) in the driving environment, and how drivers can use this awareness to predict hazardous situations.
- We suggest that situational judgement is seen as distinct from situational awareness and hazard perception as it goes beyond mere awareness of hazards, and includes attitudes and motivations.
- Commentary driving is defined in this section as a method of probing the contents of awareness during driving, and it is suggested that it can either be concurrent with the driving task, or delayed.


2.1 In this section, definitions of four important concepts within the context of driving are provided. Firstly situational awareness and hazard perception are defined, since these are concepts that have already been applied to the driving task. Then, situational judgement and commentary driving are defined. Situational judgement can be thought of as distinct from situational awareness and hazard perception since unlike these two it can involve the application of attitudes to the assessment of risk in traffic situations, and it taps into knowledge of the selection of appropriate behaviours. Further, commentary driving can be thought of as an informal method of verbalising of the contents of awareness regarding driving.

### What is situational awareness?

2.2 The most commonly accepted definitions of situational awareness[1] (SA) include both perception and understanding of the current situation and the use of this information to predict what is going to happen (i.e. so that judgements can be made about what to do). The most widely accepted formal definition is proposed by Endsley (1988), and defines SA as 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future'. The 'perception' of elements in Endsley's model is referred to as 'level 1' SA; the 'comprehension of their meaning' is 'level 2' SA; and the 'projection of their status in the near future' is 'level 3' SA. In an example driving context, level 1 SA might refer to a driver perceiving a parked lorry beside a pedestrian crossing, level 2 SA would refer to the driver having an understanding that the lorry is blocking the view of the crossing, and level 3 SA would refer to the driver predicting that a pedestrian might emerge from behind the lorry.

[1] In the literature, 'situational awareness' is usually truncated to the abbreviation 'SA'. Also, the term 'situation awareness' is used synonymously with 'situational awareness'—the former is used in the UK, while the latter originates from the USA.

2.3 Slightly different uses of the term restrict SA to only referring about perception of the physical elements in the environment, with the remaining two levels of Endsley's model being referred to as situational understanding or sensemaking (Marsh, 2000). SA tends to be measured in domains that place an emphasis on maximal performance in the skill of keeping relevant task elements in awareness, rather than on the decision and action-selection processes, beliefs, and attitudes that may also impact on performance[2]. In the driving context then, models of SA would tend to be focused on what Evans (1991) has termed 'performance', rather than 'behaviour', what Elander, West and French (1993) have termed 'driving skill' rather than 'driving style, and on 'maximal' rather than 'typical' behaviour (Baughan and Keskinen, 2005).

2.4 The current report uses the most widely accepted model of SA (Endsley, 1988), and will use the following definition of situational awareness in driving:

*"Situational awareness in driving is the perception of other road users and features of the road environment, the comprehension of their meaning in isolation and/or in combination, and the prediction of their status in the near future."*

## What is hazard perception?

2.5 There are a number of different definitions of hazard perception in the literature, and no single definition seems to be accepted by all researchers. It is generally agreed that hazard perception refers to the ability to identify potentially dangerous traffic situations (e.g. Quimby et al., 1986; McKenna and Crick, 1991), and Horswill and McKenna (2004) suggest that it can be considered to be "…situation awareness for dangerous situations in the traffic environment". Most definitions are constrained in the same way that definitions of SA are: they do not consider anything beyond the mental model held of a particular hazard or hazards.

2.6 In the current context, a narrow definition, in line with that offered by Quimby et al. (1986) and McKenna and Crick (1991) is preferable, since the focus of the review is on the suitability of new methods (situational judgment testing and commentary driving) for GB driver training and assessment and a narrow definition closely mirrors the way in which hazard perception testing is administered in the GB driving theory test. The testing only looks at the time it takes learner drivers to respond to developing hazards via a button press—it does not require any response in terms of action selection, or any demonstration of an ability to carry out the appropriate manoeuvre in a car.

2.7 There are a number of reasons why a narrower definition of hazard perception may be more appropriate in the context of this review. Firstly, this 'version' of hazard perception skill (reaction time to developing hazards) has been shown to be related to accident risk (e.g. McKenna and Horswill, 1999; Wells et al., in press), while such a link is less clear for action selection. Secondly, the definition of situational judgment (see paragraphs 2.9–2.16) naturally includes aspects of behavioural response selection, as suggested by evidence in the literature (see Section 3). Thirdly, the actual implementation of some vehicle control skills in

[2] This focus on maximal performance may be due to the fact that the formal study of SA began in the military domain of fast-jet pilot air combat, meaning that the protagonists involved are almost always focused entirely on maintaining maximal SA in life or death situations.

training has been shown, if anything, to have a negative impact on road safety (e.g. Katila, Keskinen, Hatakka, and Laapotti, 1996)[3].

2.8 With this in mind, in the current report the following definition of hazard perception in driving is used:

*"Hazard perception in driving is the ability to identify potentially dangerous traffic situations as early as possible."*

## What is situational judgement?

2.9 An agreed definition of 'situational judgement' is difficult to find in the employment and selection literature—the main literature that makes use of situational judgement tests (SJTs) and interviews. Rather, it is SJTs themselves that have been defined, in terms of what they measure. Broad consensus in the literature (e.g. McDaniel and Whetzel, 2005; Creighton and Scott, 2006) seems to be that SJTs should not be seen as measuring a specific cognitive ability, although this has been argued by some authors (e.g. Sternberg et al., 2000). Rather SJTs should be seen as measurement tools that assess multiple constructs, all of which contribute to the prediction of various aspects of job performance.

2.10 Specifically, there is general agreement in the literature that SJTs measure aspects of either job-related knowledge, or job-related behavioural tendencies (depending on how they are worded—see paragraphs 3.20–3.24), and that these correlate well with general cognitive ability (*'g'*), and personality (specifically the constructs of Conscientiousness, Agreeableness, and Emotional Stability)[4] respectively (see e.g. McDaniel, Whetzel, Hartman, Nguyen and Grubb, 2006; McDaniel et al, 2007).

2.11 On the basis of these findings, the following definition of situational judgement in driving is used in the current report:

"Situational judgement in driving is the application of driving-related knowledge and behavioural tendencies (e.g. personality, attitudes, beliefs etc.) to the assessment of traffic situations, including knowledge of the appropriate driving behaviours in those situations."

2.12 There are three advantages of this definition:

1. it does not commit to exactly what types of knowledge and/or behavioural tendencies contribute to situational judgement in this context—something that requires further research in the driving domain to define;
2. it is distinct from the definitions of the related concepts of situational awareness (see paragraphs 2.2–2.4) and hazard perception (see paragraphs 2.5–2.8), because it includes factors such as beliefs and attitudes, and it covers the selection of responses to given traffic situations; and
3. it is logically and intuitively compatible with the notion that drivers will vary in how 'good' or 'bad' their situational judgement is.

[3] It has been pointed out in the literature (e.g. Glad, 1988, cited in CIECA, 2002) that this is probably only the case when courses, such as compulsory skid training in Norway, are taught with a focus on using the control skills themselves rather than avoiding situations when they may be needed.

[4] General cognitive ability is essentially intelligence. The three personality constructs mentioned here are factors of the widely-accepted 'Big Five' personality model of Costa and McCrae (1992, cited in Cooper, 2002). Cooper, 2002 provides a readable introduction to both of these.

2.13 This third advantage is important, even if situational judgement tests are not part of a pass/fail decision in driving assessment (i.e. if they are only used to draw the attention of drivers to key risks and hazards in particular traffic situations). Put simply, 'good' situational judgement is defined as a correct assessment of the risks and hazards involved in a traffic situation, based on knowledge and/or relevant behavioural tendencies. 'Bad' situational judgement is the opposite.

2.14 The coverage of both 'knowledge' and 'behavioural tendencies' will be crucial if SJTs are to be implemented formally, since it is possible for a driver to have all the knowledge needed to assess a traffic situation correctly (e.g. "This parked lorry is a hazard as it blocks my view, and there may be pedestrians behind it about to cross the road"), but still to possess bad situational judgement due to incorrect behavioural tendencies (e.g. "I am going to drive past the lorry at my normal speed, as I am confident I will be able to brake if any pedestrians step out as I have really quick reactions").

2.15 The problem of a mismatch between the demands of the driving task and the perception of one's own driving skills has been investigated by Kuiken and Twisk (2001) among others. The ability to match one's skill to the demands of the driving situation is referred to as 'calibration'. Calibration has predominantly been investigated with young novice drivers (e.g. Finn and Bragg, 1986; Groeger and Brown, 1989) to account for their over-representation in accident statistics. In these studies, perceived capability and perceived task difficulty have typically been compared between young (18–25 years) and older drivers (35–50) drivers groups (Matthews and Moran, 1986). Calibration is discussed in more detail in paragraphs 3.43–3.47)

2.16 The wording of items on SJTs will be crucial to allow driving instructors and examiners to assess both knowledge and behavioural tendencies (see Section 3).

## What is commentary driving?

2.17 As with situational judgement, there does not seem to be an agreed definition of commentary driving in the literature. This is perhaps due to its 'intuitive' nature—it is simply talking about what one is aware of, while driving. There are some distinctions drawn in the literature regarding the level of detail included in commentary. For example, Crundall and Underwood (1997) discuss the criticisms raised by Hughes and Cole (1986) of 'continuous report', used by Renge (1980, cited in Hughes and Cole, 1986). In continuous report, an ongoing verbal protocol, without pauses, is encouraged. Hughes and Cole (1986) prefer a more succinct method by which drivers simply say what they happen to be attending to, arguing that this is less disruptive to driving performance.

2.18 In the current review, the following 'general' definition (that does not commit to one particular style of commentary) is adopted:

*"The verbalisation of at least some of the driving-related contents of awareness, while actually driving through a situation".*

2.19 Thus, commentary driving can be thought of as a measure of SA (see paragraphs 3.56–3.65), and possibly even a way of measuring situational judgement (if verbalisation is allowed to stray into statement of appropriate attitudes and self-monitoring) that happens concurrently with the behaviour (driving) under scrutiny.

2.20 Another related concept is that of verbalisation of driving-related contents of awareness just after encountering a driving situation. For example, a driver may be asked to verbalise the risks and hazards he or she noticed earlier when driving through a roundabout. This as referred to here as 'delayed commentary driving[5]', and defined as:

*"The verbalisation of at least some of the driving-related contents of awareness after driving through a situation."*

2.21 The reader will note that both of these definitions are distinct from the definition of situational judgement given in paragraphs 2.9–2.16, since they are definitions of methods, rather than of processes. It could be argued that commentary driving and delayed commentary driving are informal methods of testing SA, hazard perception, and situational judgement.

[5] The word 'delayed' implies that the commentary should occur after the driving event. However, an anticipatory commentary could also precede a driving situation. It is also possible to conceive of other ways in which commentary could be applied to driving without it being concurrent with actual vehicle control. For example, commenting while someone else is driving, and commenting on video footage of a drive are both techniques that have been used in studies into commentary driving (see Section 4).

# 3 Evidence relevant to the implementation of situational judgement testing—from the literatures on situational judgement, situational awareness, and hazard perception

**Section 3 findings**

- In the employment selection literature, situational judgement testing has been shown to be better at predicting job performance than testing based on cognitive ability or 'intelligence' alone.
- This seems to be because situational judgement tests assess knowledge needed to do a job, and also the behavioural tendencies and personality factors needed to succeed.
- Formal situational judgement testing requires carefully designed tests, and the method for doing this is described; the way in which test items are worded is especially important, since it can alter what the test is measuring (knowledge, or behavioural tendencies).

**Section 3 recommendations**

- Formal tests of situational judgement are unlikely to be suitable for use in practical training and assessment, but would benefit theory testing.
- Implementation of situational judgment testing in practical driving training and assessment may require less formal 'questioning' or 'commentary' techniques, which are discussed in Section 4.
- Whichever method is used to test for situational judgement, it should include some testing of higher levels of the GDE matrix such as attitudes, as well as hazard perception.

3.1 This section reviews evidence for the effectiveness of situational judgement measurement techniques in personnel selection for employment. The main focus is on formal situational judgement tests (SJTs). Evidence for the advantages of SJTs over and above traditional selection techniques such as interviews and intelligence tests is examined, and key features of SJTs relevant to their correct implementation are reviewed. Evidence for their use in driver training and assessment is then examined. Finally, the related concepts of situational awareness and hazard perception in driving are explored, with a focus on findings and lessons learned from these literatures that may have relevance for the implementation of situational judgement testing.

3.2 Despite the lack of evidence directly from the driving domain, SJTs are suggested as being highly suitable for use in driver training and assessment, as long as attention is paid to their correct design and implementation. It is also suggested

that SJTs should be used to focus predominantly on measuring elements from higher levels of the GDE matrix (especially aspects of self-evaluation) with elements from lower levels of the matrix (especially knowledge and skills) already being covered by the testing of hazard perception (effectively situational awareness for road hazards).

3.3 Importantly, this section aims to show that situational judgement itself (as defined in paragraphs 2.9–2.16) is best measured by situational judgement tests (as it is in the employment and selection literature). It is acknowledged, however, that full implementation of SJTs would probably only be feasible in theory testing (due to time constraints), but that a stripped-down version of situational judgement testing may be plausible and useful in practical driver training and assessment—this stripped-down version would probably look like delayed commentary driving (see Section 4), and so would need to take account of a number of caveats presented in that section, and in the recommendations in Section 5.

## Situational judgement tests (SJTs) and interviews in personnel selection

3.4 As discussed in paragraphs 1.12–1.14, situational judgement tests are defined in the literature as tests that present job-related situations to candidates, and ask candidates to select either the most appropriate behavioural responses to those situations, or the behavioural responses that candidates themselves would carry out. Some additional examples of SJT items from the literature are presented in Appendix C. In this section evidence for the effectiveness of situational judgement techniques in the largest relevant literature (personnel selection and employment) is examined.

3.5 One of the aims in the employment assessment literature is to identify variables, e.g. psychological constructs such as conscientiousness or extraversion, that accurately predict job-relevant outcomes such as job-performance or income. Identifying valid predictors of job-performance helps improving the selection of successful candidates for jobs. Different psychological variables, such a personality factors, cognitive ability or situational judgement performance are typically compared in terms of how well each of them predicts the outcome variable, e.g. job-performance. The quality of a variable as a predictor for a particular outcome such as job performance is referred to as validity. Predictors can be assessed in terms of how much of their predictive power is unique to them or how much of it they may share with other, related variables. The degree to which two variables share predictive power is indicated by the so-called incremental validity. When assessing two predictor variables, the incremental validity shows how much the prediction of the outcome measure is improved by the second variable, once the first variable has been taken into account.

3.6 Overall, the evidence for the effectiveness of situation-based measures for the prediction of job performance is strong. Situation-based measures of job performance tend to be well accepted by candidates (e.g. McDaniel and Whetzel, 2005). Situation-based measures also predict job performance better than purely psychological measures of constructs such as cognitive ability, personality variables, and even combinations of these (e.g. Chan and Schmitt, 2002, Clevenger et al., 2001, O'Connell et al., 2002, and Weekly and Jones, 1997, 1999, all cited in McDaniel and Whetzel, 2005a; McDaniel, Whetzel, Schmidt, and Maurer, 1994; McDaniel, Hartman, Whetzel and Grubb, 2007).

3.7 The available evidence for the effectiveness of SJTs in selection and employment centres mainly on the question how well they predict job performance as an

outcome measure. The following paragraphs show how the validity of SJTs as predictors manifests itself and the implications this may have for the implementation of SJTs in driver training and assessment are considered.

3.8 One of the key reasons for the success of SJTs is that they have been shown to have incremental validity[6] over general cognitive ability (*g*) in a number of studies (e.g. Chan and Schmitt, 2002; Clevenger et al., 2001; O'Connell et al., 2002; Weekly and Jones, 1997, 1999, all cited in McDaniel and Whetzel, 2005). This means that once cognitive ability is taken into account, SJTs further improve the prediction of job performance. It should however be noted that the increment above *g* is usually rather small (in the order of 2%–4% extra variance explained). SJT's incremental validity over *g* could be due the fact that as measurement methods, SJTs seem to measure factors associated with personality constructs which themselves have some predictive validity above and beyond *g*[7].

3.9 Specifically, there is agreement in the literature that SJTs measure aspects of either job-related knowledge, or job-related behavioural tendencies, and that these correlate well with general cognitive ability (*g)*, and personality (specifically the constructs of Conscientiousness, Agreeableness, and Emotional Stability) respectively (see e.g. McDaniel, Whetzel, Hartman, Nguyen and Grubb, 2006; McDaniel et al, 2007).

3.10 The extent to which particular types of SJT predict job performance seems to depend on the type of job being selected for. Evidence for this is discussed by O'Connell, Hartman, McDaniel, Grubb and Lawrence (2007). Those authors point out that job performance is typically thought of as consisting of two dimensions: task performance; and contextual performance (Borman and Motowidlo, 1993; cited in O'Connell et al., 2007). While task performance refers to expected work behaviours and actions required to perform a given job successfully, contextual performance refers to 'softer' or 'citizenship' behaviours—behaviours that benefit an organisation but are not necessarily dictated by job requirements.

3.11 A number of studies have shown that cognitive ability (*g*) tends to be the best predictor of task performance, while non-cognitive constructs such as Conscientiousness add incremental validity above *g* to contextual aspects of job performance only (e.g. Borman and Motowidlo, 1993; Hattrup, O'Connell, and Wingate, 1998; both cited in O'Connell et al., 2007). It seems likely that this is the reason why SJTs only show incremental validity above *g* to contextual elements of job performance; they measure these non-cognitive or 'softer' constructs (see however Chan and Schmitt, 2002; cited in O'Connell et al., 2007). O'Connell et al. (2007) explored this question and found evidence supporting the position that SJT predominantly measure contextual elements of performance. For jobs low in cognitive complexity (i.e. jobs likely to require 'softer' or 'citizenship' behaviours) SJTs did add a small amount of incremental validity for task performance over ability tests (correlation with job performance rose from 0.15 to 0.18). They, however, added almost nothing over ability tests and personality measures combined (0.20 to 0.21).

3.12 Another desirable feature of SJTs is that they have been shown to be fairer to sub-groups of the population than tests of cognitive ability (*g*). SJTs show fewer differences between sub-groups in a sample, e.g. between different ethnic groups or different genders, than tests of *g*. For example, O'Connell et al. (2007) examined the degree of racial and gender differences on SJTs and tests of *g*, as

[6] Put simply, incremental validity is the extra predictive power possessed by a measure, over and above another measure known to predict an outcome variable—in this case, job performance.
[7] It should be noted that *g* remains the best individual-construct predictor of job performance (McDaniel et al., 2006; McDaniel et al., 2007).

well as tests various personality variables. They found that the standard mean difference ('d'—Cohen, 1977; cited in O'Connell et al., 2007) between white and black subgroups was d=0.38 for SJTs, but d=0.66 for cognitive ability tests. The standard mean difference (d) is the difference between the two groups in standard deviation units, so a d of 0.66 indicates that one group was scoring 0.66 of a standard deviation more than the other group. The positive values of d here indicate that the tests favoured white participants (i.e. white participants achieved higher scores). Values between d=-0.01 and d=0.17 were observed for personality variables. For gender differences values were d=-0.27 and d=0.30 for SJTs and ability tests respectively, a positive score indicating that the tests favoured males over females. Differences ranging from d=-0.24 to d=0.11 were observed for personality variables.

3.13 The equal treatment of minority sub-groups is something that is important if such tests are to be seen as a 'fair' assessment of the constructs they measure, rather than simply biased towards particular groups due to the way in which they are worded (Creighton and Scott, 2006; O'Connell et al., 2007). It is also worth noting that Creighton and Scott (2006) suggest that no meaningful differences have been found on SJTs for different age groups or for respondents with disabilities. This is of particular relevance when considering their use in driver training and assessment.

3.14 Findings from the employment and assessment literature suggest that SJTs seem to be measurement methods that are acceptable to candidates (i.e. that possess face validity), and that can be used to assess *g*, and a number of personality constructs, which together seem to be good predictors of the task and contextual dimensions of job performance. They also give rise to fewer sub-group differences than tests of cognitive ability. Situational-based measures such as SJTs therefore present a viable method of measuring relevant knowledge and behavioural tendencies during driving training and assessment, especially if the aim is to include higher levels of the GDE matrix in the driver training regimen.

## Key features in the implementation of SJTs

3.15 This section considers the implementation of SJTs for the purpose of measuring situational judgement. Particular emphasis is placed on the way in which items (and answers) are constructed, and on the way in which response instructions are worded. Both of these issues have profound implications for the way in which SJTs may (or may not) be implemented as part of the GB driver training and assessment.

### *Construction*

3.16 McDaniel and Whetzel (2005b) suggest that there is no 'rule book' for the construction of SJTs, except that SJTs consist of a 'stem' (e.g. "Everyone in your work-group has received a new computer, except you"), and several item responses (see Appendix C for responses to this stem). However, McDaniel and Whetzel (2005b) also claim that there are eight characteristics on which SJTs tend to differ. These are:

1. **Fidelity of stems.** This is the extent to which the SJT stem format represents the exact situation. High fidelity would, for example, be a video, while lower fidelity would be written form.
2. **Stem Length.** The length of stems differs greatly between different SJTs.

3. **Stem complexity.** The complexity of the situation described.
4. **Stem comprehensibility.** The degree to which the stems are easy to understand in terms of their readability.
5. **Nested stems (or not).** Some tests have a description of an overall situation, and then several 'nested' stems related to different problems that need to be dealt with within this overall situation.
6. **Nature of responses.** Usually responses are short and written.
7. **Nature of response instructions.** Typically seen to fall into two categories—focused on knowledge, and focused on behavioural tendencies (see paragraphs 3.20–3.24)
8. **Degree of item heterogeneity.** Typically SJTs measure multiple constructs, but this can vary.

3.17 The proposed characteristics of SJTs imply that tests can vary hugely. Even if there were only two different discrete versions of each of the eight variables above, this would mean that it would be possible to have 256 (i.e. $2^8$) different 'types' of SJT[8]. This variability probably goes some way to explaining why McDaniel and Whetzel (2005b), among others in the literature, recommend thinking of SJTs as a measurement method which can measure multiple constructs.

3.18 In terms of how the items and stems for SJTs are developed, again McDaniel and Whetzel (2005b) provide an overview that illustrates the variability inherent in these measurement tools. The authors suggest that these stages are usually followed in development of an SJT:

1. **Write critical incidents for the job to be assessed.** The authors suggest setting up a group of subject-matter experts (SMEs) to write the critical incidents, and an additional set of SMEs to write the responses. This process can be facilitated through having workshops, and prompts to encourage the SMEs to think of, for example, particular times when someone performed very well on the job in question. McDaniel and Whetzel (2005b) give several recommendations for such prompts. If implemented for the driving task, the SMEs would be driving examiners, instructors etc., and the critical incidents would be hazardous driving situations.
2. **Sort the critical incidents.** In this stage, the critical incidents need to be sorted into categories. For example in driving, there may be categories linked to pedestrians, other traffic, un-sighted hazards, surface hazards etc. At this stage, duplicate incidents can be discarded, and the remaining incidents can be turned into stems (next stage).
3. **Turn selected critical incidents into stems.** This is extremely labour-intensive, and needs to be done in such a way that the stem is relevant for all people who will be asked to complete the SJT item. This should be easier in the driving domain than in employment, since the driving task is more homogenous than general employment. For example, in all driving, a situation judgement stem about child pedestrian is relevant, while a situation judgement stem about computer software is only relevant for jobs involving software. The importance of stem clarity and brevity are of considerable importance.

[8] In reality, even though McDaniel and Whetzel (2005b) point out that stem length, complexity and comprehensibility are highly correlated, the real number of different 'types' of SJT will be even higher since clearly each of these variables will give rise to more than two 'versions' (i.e. video, audio, and written are at least three levels of fidelity, in decreasing order of the realism with which they represent the situation being judged).

4. **Generate and edit item responses.** After stems are written, lists of alternative responses need to be generated. Again this is labour intensive. Responses should be written by SMEs with varying levels of ability in the domain, and there should be a variety of good and bad responses for each stem. The authors suggest that multiple SMEs will be needed to ensure that enough non-redundant responses for each stem are generated (with individual SMEs only being able to write 2 or 3 each). Responses that will have almost no variance in answers should also be avoided.
5. **Generate response instructions.** The decision needs to be made to go with either 'knowledge-based' or 'behavioural tendency' response instructions (see paragraphs 3.20–3.24). This has profound implications for what the test is likely to measure.

3.19 It is outside of the scope of the current review to go into detail regarding the later stages (developing scoring keys; validation studies) that are required to finally develop and validate formal SJTs (for driving training and assessment purposes). The important point is that SJT development is a rigorous scientific process that needs to be taken seriously. Any future implementation of formal SJTs in driver training or assessment will need to take these procedures into account, especially if situational judgement 'scores' are to be used in assessment as a 'pass/fail' criterion.

### *Wording of response instructions*

3.20 Something related to the extent to which SJTs have incremental validity is the way in which response instructions are worded. Response instructions can be worded with respect to knowledge (e.g. "which is the most appropriate response from those below?"), or with respect to behavioural tendencies (e.g. "which of the following responses would you be most likely to perform?"). McDaniel and Whetzel (2005a) suggest that this can influence how much an SJT measures *g* (knowledge) or personality (behavioural tendencies), and thus how much incremental validity above *g* it is likely to have. There is evidence for this suggestion—for example, McDaniel, Hartmann and Grub (2003) showed that if response instructions were worded with respect to knowledge then scores on an SJT were correlated most highly with cognitive ability, while if the instructions were worded with respect to behavioural tendencies, then scores on SJTs were correlated most highly with personality variables such as Conscientiousness, Agreeableness, and Emotional Stability.

3.21 Response instructions will also impact on how easy an SJT is to fake, with various authors (e.g. McDaniel and Whetzel, 2005b; McDaniel et al., 2006) suggesting that response instructions worded with respect to behavioural tendencies will be easiest to fake. Social psychology studies have shown that self-reports of personality, attitudes and behaviour are frequently inaccurate or biased to some degree, because at least some participants tend to engage in socially desirable responding (SDR), i.e. a tendency to give answers that make the respondent look good (Nederhof, 1984; Paulhus, 1991; Paulhus and Reid, 1991).

3.22 Two factors have been identified to contribute to SDR: so called "impression management", the deliberate tendency to give favourable self-descriptions to others (Paulhus, 1991; Paulhus and Reid, 1991; Verkasalo and Lindeman, 1994); and "self-deception", a personality construct that is linked to psychological adjustment (Sackeim and Gur, 1979; Taylor and Brown, 1988), high self-esteem (Paulhus and Reid, 1991), and lack of neuroticism (Borkenau and Ostendorf, 1989). Self-deception is thought to provide an aid for coping with negative life events and threatening information (Paulhus, 1984; Paulhus and Reid, 1991).

Self-deception plays a particular role in the development of over-confidence in one's own capabilities as a drivers and is thus of considerable importance in the field of learner drivers. Despite its obvious importance, research on the topic is scarce (Lajunen, Corry, Summala and Hartley, 1997).

3.23 Considering the potential impact of SDR and ways of controlling for it will be an important requirement in the inclusion of situational judgement testing in driver training and assessment. Psychometric scales to assess the extent of self-deception and impression management in the driving context have for example been developed by Lajunen et al. (1997) and may be considered as a tool for controlling SDR in situational judgement testing. Some authors such as McDaniel et al. (2006) claim that it is unlikely that the validity of situational judgement testing as a whole will be undermined as a result of faking. These authors however concede that further research is needed to assess the magnitude of the problem of faking.

3.24 Sub-group differences have also been shown to be sensitive to response instructions. Nguyen and McDaniel (2003) showed that two versions of the same SJT with knowledge instructions and behavioural tendency instructions both showed black-white differences, but that the knowledge instructions (due to the higher cognitive saturation of the test—the extent to which it appeared to measure *g* rather than personality variables) gave rise to larger ethnic sub-group differences. Again this consideration is highly relevant to driver training and assessment, since it is crucial that any training and assessment methods used are best placed to avoid unfairly disadvantaging sub-groups.

## Evidence for the effectiveness of SJTs in driving

3.25 In addition to the use of situational judgement in the employment selection literature, attention is now turned to the use of situational judgement-type measures in driver training and assessment. Again the focus is on reviewing evidence that has relevance to the implementation of situational judgement testing in Great Britain's driver training and assessment.

### *European projects*

3.26 The GDE Matrix introduced higher cognitive levels into existing hierarchies of skills for the driving task (Hatakka et al., 2002). Several authors on a number of key European projects have since been pushing for including these higher-level skills into training and testing. The implementation of some form of situational judgement testing fits well into this programme of change. Brief overviews on relevant European projects are provided in the following.

### *The TEST project*

3.27 The TEST project (Baughan, Gregersen, Hendrix, and Keskinen, 2005) examined driver testing in six European countries. One of the objectives of TEST was to examine whether tests assess the relevant skills, behaviours and knowledge in learner drivers and to develop recommendations to improve testing where necessary. The final report mentions the gradual blurring of the distinction between practical and theory elements of the driving test, as more use is made of novel training and assessment techniques to train and assess different elements of driving skill. This is particularly important when considering the aim that training should comprehensively cover all the skills and knowldege that novice

drivers require to be safe, rather than relying on the test to be a 'filter' for safe drivers and unsafe drivers.

3.28 The TEST report recommends the use of self-report as a way of trying to understand what controls observed behaviour: the driver can be asked what he thinks he is doing, and why he is doing it. The TEST report suggests that this is an attractive and flexible approach, and well-worth pursuing; though it is not without its own difficulties, including:

- People do not always have good insight into the reasons for their behaviour and hence may not be able to report them correctly, even if they wish to (see paragraphs 4.36–4.47).
- The technique is open to bias in that drivers may know, and report, the desirable answer, just as they may be exhibiting desirable (maximal) rather than typical driving behaviour. As shown in paragraphs 3.20–3.24, the wording of test items can be used, to some degree, to offset this.
- Asking for verbal reports may change behaviour. For example, by reminding a candidate that certain types of behaviour are good, or by increasing workload (see paragraphs 4.2–4.9).
- People who find it difficult to express themselves verbally will tend to be at a disadvantage. See paragraphs 4.49–4.51 for evidence that this is the case even with advanced Police drivers using commentary driving.

3.29 The point is then made in the TEST report that in situations where we cannot expect to have typical behaviour to observe, an alternative is to discover whether the candidate has (a) knowledge and insight into goals, personal characteristics and risk increasing factors in everyday driving situations, (b) knowledge of how these factors can affect driving and (c) knowledge and skills to enable him to make good, safe, decisions in the light of (a) and (b). In other words, if a candidate's typical behaviour cannot be observed directly, we can at least try to discover whether he has the prerequisites for 'acceptable' typical behaviour.

3.30 The testing of levels three and four of the GDE matrix[9], it is suggested, might involve the following:

- Knowledge tests covering influences and risks associated with life goals, social pressures, behavioural style, substance abuse, on driving.
- Tests of knowledge about how such influences can be dealt with to maintain driving safety.
- Tests of decision-making behaviour in which scenarios are presented, and questions are asked.

[9] Another area for consideration in the use of situational judgement testing as an on-road intervention in training is the use of anticipated negative emotional consequences. Richard, van der Pligt and de Vries (1996, cited in Engstroem et al., 2003) showed that education regarding the negative consequences of catching a sexually transmitted disease only produced behavioural change if it was framed in such a way to promote imagining of the negative emotions involved ('imagine the feeling of waking up having had unprotected sex with an occasional partner' versus 'imagine the long term medical effects of this disease'). It may be that the changing of higher level cognitions and attitudes in the higher parts of the GDE matrix will only work through such emotional mechanisms, rather than through assessment of 'road sense' in specific traffic situations (which is still useful, but may have an effect through a similar mechanism as that used by the hazard perception test—the mastery of traffic situations). Such an intervention in UK driver training would probably work best in group discussion settings.

- Extended multi-media based 'hazard perception' tests in which scenarios involving various risks are presented.
- Self-evaluations, and comparisons of these with evaluations of examiners (and/or possibly peers, instructors and supervisors).

3.31 It is also acknowledged in the TEST report that although pass-fail tests for some of these are in principle straightforward to devise (although see paragraphs 3.16–3.19 for some of the issue associated with formal SJTs), for others there will be difficulties. For example:

- An individual's self report of goals, attitudes, and beliefs are open to bias.
- Defining pass/fail criteria is not straightforward. For example, what standards are good enough? This is made more difficult by the fact that there will be large variation between candidates in terms of 'goals for life and skills for living' whereas the test needs to select people not on this basis, but on the basis of how well they are able to recognise and cope with these personal characteristics and goals, and how they are able to match these characteristics to driving situations—what Kuiken and Twisk (2001) call 'calibration'.
- Making the assessments may be difficult—it will be important to train examiners and instructors appropriately.
- There may be difficulties in explaining and justifying the assessments and decisions to candidates, instructors and (occasionally) in response to legal challenge. It is noted that different countries have different traditions and legal requirements.
- Traditions in legal systems, testing practice, and instructor and examiner selection and training may make it difficult to accept a change from objective records of errors to other, less tangible, pass/fail criteria.

3.32 The authors of the TEST report recommended that even if pass-fail criteria are not plausible, it is of benefit to cover levels three and four of GDE in training and assessment: not to contribute to a pass/fail decision, but to draw attention to the importance of these goals and attitudes in contributing to safe behaviour.

3.33 A final point made in the TEST project was that current computer-based hazard perception testing could be expanded to include levels three and four of the GDE matrix. This suggestion may apply to the wider context of driver assessment and training and thus to situational judgement testing—there is no reason why techniques as situational judgement testing cannot be built more into the theory test, since this will permit greater control over the situations seen by candidates, and will avoid some of the problems identified in this report with using such techniques 'live' on the road.

### *The HERMES project*

3.34 Support for situational judgement type-testing also comes from the HERMES project (CIECA, 2007). This project aimed to create a short (3–5 days) training course to help driving instructors enhance their 'coaching' skills. This is important, especially in light of the efforts being made in various countries to improve coverage of the higher levels of the GDE matrix in driver training and assessment. Bartl observes in the HERMES report that the challenge of coaching "is to lead the student out of the role of a passive consumer and into the role of an active producer" (Bartl, cited in CIECA, 2007, p6). The aspiration to include

situational judgement approaches and commentary driving in Great Britain's driver training and assessment can be seen as being commensurate with this goal. By encouraging learner drivers to explore their own thought processes when driving in specific traffic situations, learner drivers can become more aware of their own responsibilities, and can engage some of the higher level thought processes described in the GDE matrix. Importantly, by enabling students to engage with their own learning process, good educational and learning practice is ensured (e.g. Holt, 1967, cited in CIECA, 2007). An important implication of the potential inclusion of situational judgment testing or commentary driving in training and assessment is that driving instructors and examiners may need additional training, both on the correct administration of the tests, and in coaching methods that are pertinent to more open-ended types of testing for situational judgement. If the aim is to access higher cognitive processes that occur in the learner when negotiating his/her route through the dynamic traffic environment, question and answer techniques and coaching would represent important skills in driving instructors. Therefore, deciding on the format of the situational judgment testing that is to be implemented, account will need to be taken of the training needs of examiners and instructors.

### *Does situational judgement testing belong in the theory test?*

3.35 As has been stated already, some authors (e.g. Baughan et al, 2005) have suggested that a consequence of aiming to test the higher levels of the GDE matrix is the potential blurring of the distinction between the theory and practical elements of driver assessment. The conclusion from this review is that the theory test (or possibly a taught course—see paragraphs 3.36–3.38) seems the most appropriate place within the assessment protocol to include formal SJTs. The degree of standardisation and rigorous development of SJTs may not permit their use during practical testing. In this case, the only option open for examiners and instructors 'on the road' to getting at the contents of a driver's awareness is some form of commentary (or delayed-commentary) driving (see Section 4).

### *Coaching skills*

3.36 Another implication of the move from lower levels to higher levels of the GDE matrix in driver assessment is that the trainers may themselves require training in new skills associated with 'coaching' drivers on attitudes. These skills have been identified as being very different to the skills required to teach car control skills (CIECA, 2007).

3.37 Degia (2007) reviewed the literature on skills necessary for good 'coaches'. From this review, Degia discussed three aspects of the qualities required by good coaches: personal attributes (i.e. things that are unlikely to be teachable); skills that can be taught; and practical requirements that coaches need to possess. Degia concluded that coaches should possess the following personal attributes: an ability to 'get on' with young people (since young people make up the majority of new drivers); good interpersonal skills; and an ability to be patient regarding mistakes. Skills to be taught to coaches include: avoiding coming across as an 'expert'; a calm and relaxed attitude to coaching; encouraging candidates to develop self-analytic techniques; use of participant-centred techniques; being able to deliver clear reinforced messages that do not alienate individual participants; and being able to motivate. Finally, practical requirements needed to ensure that coaches come across as credible to their candidates include: the ability to model positive behaviour; good driving skills; 7+ years driving

experience; being aged 25 years or more, with a clean driving license; and able to satisfy a police check.

3.38 Degia points out that most of the skills needed can clearly be taught to those existing ADIs who possess the correct personal attributes, while the practical requirements are very similar to those already required for ADIs. It is likely that the development of skills necessary to teach learner drivers learning goals located in the higher levels of the GDE matrix would require specific course (possibly in groups) for ADIs interested in acquiring these skills. Cascade trainings, where those ADIs that have already undergone the training teach the ones that have not, could be considered. It is unlikely that all ADIs possess the skills or motivations to perform a coaching role, and attempting to change the job specification for all ADIs may have the adverse effect of driving potential ADIs away from the career.

### *Implementation of situational judgement testing in the Netherlands*

3.39 There have been several attempts to actually implement situational judgement techniques. For example, De Craen et al. (2005) developed a post-licence training package on higher order skills in the Netherlands, and included a brief situational-judgement-type test involving speed choice estimates for static pictures of traffic scenes that differed in complexity (i.e. in a 'theory'-type setting, rather than practical 'in-car'). For example, a scene in which the driver's car is approaching a crossroads either is clear of traffic (lower complexity), or has a single cyclist apparently looking to turn across the driver's path (higher complexity). The expected result on speed choice if a driver is assessing the situation correctly is to offer a slower speed for the higher complexity scene, and this measure can also be used to assess the effect of training interventions on situational judgement.

3.40 Findings were mixed—only some participants showed a safety increase after training, and only on one of the two key scenes used. Also, control participants who received no training also improved their safety discrimination between the lower and higher complexity scenes over the same time period, while other control participants who received slightly different training got worse in terms of giving higher speed estimates to the higher complexity scenes. Clearly though, situational judgement-type questions can be tailored for the driving domain, and it is possible to define 'correct' responses for given situations, in this case in terms of differentiation of lower and higher complexity scenes.

### *Implementation of situational judgement testing in New Zealand*

3.41 In the New Zealand full-licence test, during Part 2 of the test (detecting and responding to hazards in built-up areas) candidates are asked to stop after making particular traffic manoeuvres, and are asked to describe the major hazards that they noticed when they were performing it. The hazards noticed must map onto those noted by the examiner. During Part 3 (detecting and responding to hazards in high-speed areas) the candidate is expected to describe hazards while actually performing the manoeuvre. In both cases, it is expected that the manoeuvre is carried out safely. There has not been any major evaluation of this test since the basic evaluation in its inception (Christie, personal communication to first author, 2008). However, it is clearly an example of a testing procedure where some kind of situational judgement testing procedure is included (although less formal than SJTs and more like 'delayed commentary' driving—see paragraphs 2.17–2.21).

### *Studies of situational judgement testing in the USA*

3.42 Fisher, Pollatsek and Pradhan (2006) showed that giving novice drivers PC-based training of where to look for information that may be hidden from view in a driving scene (for example, pedestrians approaching a crossing being hidden by a parked lorry) can result in drivers exhibiting increased gaze towards those areas in similar situations both in a simulator and on a real-road drive. Interestingly, performance at a PC-based task of similar driving situations as those used in training also showed the training effect, as did the simulator and on-road measures. Drivers who had not been trained were poor at indicating on a PC which areas of a plan-view driving scene should be looked at for such 'hidden risks', while drivers who had been trained in the general principle of hidden risks performed much better in this task. Again, this is an example of something that, while not strictly speaking an SJT, is a demonstration that it may be possible to evaluate important principles of situational judgement in simple, PC-based tasks.

## Situational judgement testing and improving 'calibration'

3.43 Kuiken and Twisk (2001) report that learner drivers tend to overestimate their skills and underestimate the demands of the driving task (e.g. Gregerson, 1996, cited in Kuiken and Twisk, 2001; Horswill, Waylen and Tofield, 2004)—a phenomenon that Kuiken and Twisk have termed 'poor calibration'.

3.44 An example of this is the fact that people tend to judge their own driving ability as better than that of other drivers. This 'illusory bias'[10] was demonstrated by Horswill et al. (2004), who showed that when drivers were asked to rate their own ability on a range of driving skills, relative to their peers and to the average driver, they rated themselves as having more ability than peers, or the average, on the majority of those skills. In addition, the illusory bias was greater for hazard perception skill than for vehicle control skills, which is interesting since hazard perception skill is known to relate to accident risk. If drivers are especially prone to overestimating this skill (i.e. having poor calibration) as Horswill et al's data show, then this has serious implications for road safety.

3.45 Recent work indicates that drivers also show poor calibration in terms of their awareness of performance decrements introduced by carrying out a secondary task (telephone conversation) when performing a driving task. Horrey, Lesch and Garabet (2008) had younger and older drivers perform a drive on a track, and included three performance measures: stopping time to a red light; pace-keeping performance (speed control); and lane keeping performance (lateral control). In addition, drivers were asked to rate their performance subjectively in each task. In all three tasks, the telephone conversation conditions resulted in objective performance decrements relative to baseline. However, there was no correlation between objective performance decrements and subjective performance decrements for drivers overall. Thus overall, drivers had very poor calibration—they were not aware when they were performing well, or poorly. When different sub-groups of drivers were examined separately, it was found that female drivers showed the same poor calibration as the overall group. However, older male drivers showed good calibration between objective and subjective performance

[10] The belief has been described as 'illusory' because at the group level, it is simply not possible for the majority of people to be better than the majority of people (Taylor and Brown, 1988; Weinstein, 1980; both cited in Waylen, Horswill, Alexander and McKenna, 2004). Illusory biases have been demonstrated in a number of domains, including driving.

(i.e. they reported poor performance when they performed poorly, and reported good performance when they performed well), while younger male drivers showed the opposite pattern (i.e. even worse calibration than the group overall—reporting good performance when performing poorly, and vice-versa). The fact that younger male drivers show the worst calibration is relevant for driver training tackling the well-established increased accident risk of younger, less experienced drivers, and in particular young males.

3.46 Kuiken and Twisk (2001) have suggested that one way to enhance safety is to improve calibration—to ensure that drivers (in particular, younger drivers) are aware of the mismatch that can exist between their perceived and actual skill in hazardous driving situations. Awareness of this discrepancy should enable drivers to better control their driving to ensure that task demands at any given time are in line with their abilities. If drivers might be encouraged to understand their lack of ability in a given situation, they may drive more cautiously as a result.

3.47 It is plausible that situational judgement tests may help learner drivers develop insight into their actual driving skills, and thus help them to develop better calibration, through a focus on specific driving situations.

## Hazard perception

3.48 It is generally agreed that hazard perception refers to the ability to identify potentially dangerous traffic situations (e.g. Quimby et al., 1986; McKenna and Crick, 1991). It is generally measured using video-based tests, in which participants view traffic scenes filmed from the perspective of the driver in a car. Participants are required to respond to specific developing hazards by pressing a button, or clicking on the relevant area of the screen with a mouse pointer (e.g. McKenna and Crick, 1991; McGowan and Banbury, 2004). The usual measure of hazard perception ability is the anticipation time for these specified hazards—larger anticipation time equates to better hazard perception skill, since it reflects that participants are seeing and responding to hazards earlier in their development.

3.49 Hazard perception skill is related to experience (e.g. McKenna and Crick, 1991; McKenna and Horswill, 1999), and has been shown to be related to accident risk (e.g. Quimby et al., 1986; Hull and Christie, 1993; McKenna and Horswill, 1999). It is also a skill that is trainable. For example, McKenna and Crick (1993) and Sexton (2000) have shown that when drivers are trained using video-based stimuli, and are asked to predict what might happen next in those scenes when they are frozen, then such drivers are later better at hazard perception in validated tests (not involving the same road scenes). In addition, road-based training has been shown to enhance hazard perception ability (Crick and Mckenna 1991). A recent review of the main findings from the hazard perception literature can be found in Horswill and McKenna (2004).

3.50 Hazard perception testing is already used in the GB driving theory test, on the strength of the key findings from the literature. In this section, we review some findings from the hazard perception literature, with a specific focus on what they can tell us about the suitability of situational judgement testing techniques (including SJTs) in GB driver training and assessment. Again, although there is overlap between the theoretical terms, the hazard perception literature presents some findings that also highlight the differences between the two forms of testing.

3.51 An early paper is that of Armsby, Boyle and Wright (1989). These authors examined multiple methods for testing hazard perception before video-based tests had been developed, including interviews, Q-sort, and repertory grids[11]. They concluded that interview techniques were not useful in producing drivers' perception of hazards, without visual presentation of particular road scenes. Armsby et al. pre-empt a key point in the literature related to what should be tested for—hazard perception as a skill seems to require the kind of dynamic stimuli used in video-based tests (and see Wallis and Horswill, 2007, below), while SJTs and similar measures may be best suited to measuring higher levels of the GDE matrix.

3.52 Fitzgerald and Harrison (1999) surveyed 50 driving instructors in Victoria, Australia regarding the training of learner drivers, and their attitudes towards hazard perception. Results showed that instructors were aware of the importance of hazard perception skill, with 20% of them rating it as the most important skill young drivers can possess. The instructors also recommended a number of other methods of testing hazard perception in addition to computer-based testing, including asking questions of learner drivers during driving; again supporting the notion that situational judgement testing, or rather commentary driving of some kind, may be a plausible way of testing hazard perception 'on the road'[12].

3.53 Another important finding that is directly relevant to the use of situational judgement tests comes from Wallis and Horswill (2007; see also Farrand and McKenna, 2001). Novice and experienced drivers' ratings of the *potential* hazardousness of road situations were the same when this judgement was made (un-timed) in an occlusion paradigm[13]. This finding suggests that novices and experienced drivers are both equally able to judge the level of risk for traffic conflicts present in a scene when questioned in an un-timed 'out of task' way. Importantly however, when drivers were asked to indicate when hazardous road situations were actually occurring in the standard hazard perception test format (i.e. press a response button as soon as they see an actual hazard developing), experienced drivers responded earlier to hazards than novices.

3.54 Wallis and Horswill suggested that this shows experienced drivers' advantages in hazard perception ability is based on a response bias, rather than greater sensitivity to detecting hazardous situations. Put simply, experienced drivers do not see more potential hazards than novice drivers, but they do respond more quickly to real ones because they have a lower threshold of responding 'hazard present' as dynamic cues to the occurrence of a hazard become available. This has implications for the use of certain types of situational judgement tests (e.g. 'judge the hazard risk in this road scene') in driver assessment and training, since scores on such tests may overestimate a given driver's true ability at 'reading the road' and responding to hazards as they actually drive; novice drivers seem to

[11] Repertory grids are used to measure how people impose structure on their world, by eliciting the constructs they use to define different people and objects. Q-sort is a method used to elicit people's self-perception. Cooper (2002) has an introductory chapter on the use of these two qualitative methods.

[12] These authors also introduce the notion of 'hazard behaviour' which includes the actions taken to respond to hazards as well as the perception and recognition of them. They couch 'hazard behaviour' in terms of Klein's Recognition-Primed Decision Making (RPD) model (e.g. Klein, 1989, 1993, cited in Fitzgerald and Harrison, 1999), and Logan's instance theory of automaticity (e.g. Logan, 1988, cited in Fitzgerald and Harrison, 1999). Both of these models posit that when tasks become automated with increasing practice, the mechanism by which this occurs is one of fast, automatic retrieval of memory for similar events which lead to predictable responses. Note that using such models of memory retrieval to explain the hazard perception advantages of more experienced drivers is questionable. Subsequent research has shown that even for experienced drivers hazard perception is not automatic, since hazard perception performance is interfered with if a secondary task is employed, (e.g. Rowe, 1997; McKenna and Farrand, 1999).

[13] In the occlusion paradigm, dynamic video or simulator scenes are 'blanked out' and drivers are asked to judge the potential hazardousness of the situation after the blank-out.

have the ability to judge the 'overall' hazard potential in scenes, but lack the sensitivity possessed by experienced drivers to identify hazards as hazardous early in their development. One caveat to this finding is that the novice drivers used in Wallis and Horswill's study had been driving for an average of 2.13 years. It is possible that judgements of the hazardousness of road scenes in learner drivers (as opposed to the novice drivers in Wallis and Horswill's study) will not over-estimate ability in the same way.

3.55 Soliday (1975) developed a 'hazard judgement test' that is similar in conception to SJTs. In Soliday's study, drivers completed hazard ratings on a seven-point scale for text descriptions of hypothetical driving situations that contained either a moving, non-moving, or environmental hazard (e.g. "You slow down for an intersection, but see in the rear-view mirror that the car following you is not slowing down at all and is only a couple of car lengths back", "You're on a country road in the winter and the surface is covered with a couple of inches of wet, slippery snow" or "You're coming home at night in a heavy fog on a heavily travelled freeway" respectively). It was found that drivers who reported fewer accidents over the last five years also rated environmental hazards as significantly more hazardous. This study shows that there is potential for even very simple indices of what might be referred to as situational judgement to be related to complex outcome variables such as accident risk, although it is noted in the paper that more work is needed with larger samples.

## Situational Awareness (SA)

3.56 Baughan et al (2005) recommended that the scope for strengthening the assessment of hazard perception and SA during the British practical driving test should be explored further, and that situational judgement elements should be considered for incorporation into the theory test. Similarly, a fact sheet from SWOV (SWOV, 2007), the Dutch National Road Safety Institute, suggests that hazard perception and Situation Awareness can be thought of as the same thing in the context of driving, although it does acknowledge that Situation Awareness has the advantage of formalising the definition of what aspects of a situation a driver needs to be aware of (i.e. in Endsley's model, SA goes beyond mere perception, to comprehension and prediction, although it does not have anything to say about action).

3.57 Horswill and McKenna (2004) suggest that hazard perception can be thought of as SA for hazardous traffic situations. It is easy to see why SA and hazard perception have become synonymous in some parts of the literature: for example consider the findings reported in paragraphs 3.48–3.55 from Wallis and Horswill (2007). This could be interpreted as novices and experienced drivers not differing in terms of their (in the terminology of Endsley's model, see paragraphs 2.5–2.8) level 1 SA (perception) for hazards, but experienced drivers having better level 2 SA (comprehension) and level 3 SA (projection) and thus being more likely to make a hazard response in the video-based task.

3.58 Maycock and Forsyth (1997) carried out a large study with learner drivers to investigate the links between various aspects of test performance and accident liability over the ensuing three years. They showed that 'awareness and anticipation' errors on the test were the main predictor of later accident risk, and were the only predictor in male drivers. Again these data fit with the apparent difference in level 3 SA between accident involved and accident free drivers. In McKenna and Crick's (1993) early discussion of hazard perception, they conceptualise the skill as requiring a 'mental model' that can be run forward to predict potential traffic conflicts. This suggestion has clear parallels to SA.

3.59 The allure of applying SA to driving is, therefore, understandable. SA is a well-developed theoretical construct, with (as we have seen) established models and definitions in the literature. The logic is that if the structure of SA for driving can be understood, then measures can be taken to improve road safety, and to improve the design of road systems to increase SA. This approach has been used in a number of other fields—not least the military field in which SA was originally developed, and related fields such as the control of complex systems. For example, Stanton, Chambers and Piggott (2001) discuss several studies that have sought to classify the errors made by people operating complex system control settings (e.g. aviation, air traffic control) and show that an understanding of the structure of SA can help in the process of classification, and by extension can help in identifying interventions to make such errors less likely in the future.

3.60 However, given the focus in the road safety literature on training and assessing the higher levels of the GDE matrix, it is argued here that the adoption of SA as a theoretical construct to guide research is not necessary. SA theory has been born from military contexts, in which the assumption has been made that the protagonists are always attempting to behave optimally in terms of their skill—their motivation being that the contexts in which they operate are literally 'life and death' situations. Although it is well known that driving skills such as hazard perception are indeed linked to what can be life and death outcomes (i.e. accident risk), it is generally agreed in the literature that drivers (especially novice drivers) simply do not use this exclusively as a motivating factor to guide their behaviour.

3.61 It is argued here that the only tangible way in which SA research can be applied to learner driver training and assessment is by suggesting the measurement techniques that can be used to measure SA on the road. This work has already begun, since work on hazard perception is already being integrated with theoretical work on SA, in order that the finer details of hazard perception itself can be derived (see, e.g. McGowan and Banbury, 2004).

3.62 A lot is known about measurement techniques in the SA literature, in particular the advantages and disadvantages of different techniques. Broadly speaking, there are two ways in which measures of SA differ. Firstly, they can be objective, or subjective. Objective measures of SA such as SAGAT (Situational Awareness Global Assessment Technique—see e.g. Endsley, 1995; 2000) and SPAM (Situation Present Assessment Method—Durso et al., 1998) measure whether someone has particular knowledge directly. For example, in a flight simulator, a pilot's SA for the speed and heading of an enemy aircraft could be probed directly, with either the speed and/or accuracy of the answer being used as an index of SA. Subjective measures such as 'SART (Situation Assessment Rating Technique—Taylor, 1990; Taylor and Selcon, 1991, see also Jones, 2000) simply ask people what they think their SA is, and sometimes this is broken down into different aspects of the cognitive processes involved (e.g. in SART, people are asked about how much information is being supplied to them from the environment, and also about how much mental resource they have to devote to the processing of this information).

3.63 Endsley, Selcon, Hardiman and Croft (1998) compared subjective and objective measures of SA (SART and SAGAT). These authors found that both measures were diagnostic of the SA resulting from changes in displays used during threat avoidance in a flight simulator. Those displays that increased threat avoidance also increased SA on both measures. However, the objective measure 'SAGAT' provided richer detail as to when SA was higher, and when it was actually lower (despite subjectively participants thinking it was higher). This is evidence for subjective ratings of SA being somewhat misleading—one can think that one has good SA, because one is unaware of the things one does not, objectively, know.

Although there have been some demonstrations in the literature of more subjective 'judgement' type tasks being used to measure SA objectively (e.g. Strauss and Kirlik, 2006), these rely on comparison of judgements to known system values, and therefore are not suitable for practical driving tasks, in which true values are unknown. For example, asking a learner driver to judge the likelihood of a crash in a given situation will not reveal true SA unless the likelihood of a crash is known. As a proxy for the true value in the driving context the judgement of an expert such as the driving instructor could be used.

3.64 Secondly SA measures can differ in whether they require interruption of the task to allow data to be captured. SAGAT, for example, is a so-called 'freeze' technique, whereby the task is 'frozen' to allow probe questions to be asked about various aspects of the situation. Such measures could be used in simulator-based driver training, although would be more challenging to implement on the road. Other measures of SA (e.g. SPAM) allow real-time testing. Interestingly, the GB video hazard perception test can be thought of as a continuous measure of SA (albeit using standardised 'simulation' stimuli). Again, there are advantages and disadvantages of both types of approach, and once it is certain what knowledge needs to be measured, and what practical constraints exist on collection of data, it will be possible to use what is known about different types of SA measure to guide training and assessment in learner drivers.

3.65 In summary, there are a number of interesting developments in the SA literature that will have some relevance to driver training and assessment. However, beyond a basic appreciation of the general theoretical constructs, and practical knowledge on testing techniques, there is actually little to be gained from implementing SA testing specifically in learner driver training and assessment, not least because hazard perception training already seems to be accepted as doing this.

## Summary

3.66 Situational judgement tests (SJTs) are a formal measurement technique that can be used to test for knowledge and behavioural tendencies in a given task situation. Typically, SJTs have been used in employment selection, where they have shown a number of advantages, including some incremental validity over tests of cognitive ability and personality tests. They have potential for use in driver assessment and training. Situational judgement in driving is distinct from SA and hazard perception (which themselves can be thought of as synonymous) in that to be 'good' at it, once needs to utilise safe attitudes as well as task-relevant skills. SJTs therefore offer a good way of comprehensively assessing relevant attitudes and skills.

3.67 When considering the introduction of situational judgement testing into practical driving training and assessment, more work will be needed to assess the various issues described in the current section. Specifically, formal introduction of SJTs will require a number of considerations, including: an analysis of the aspects of driving performance that need to be tested (i.e. are they 'task' or 'contextual' or both); attention to the way in which such tests need to be developed; and attention to the way in which SJT items are worded—with respect to behavioural preferences (i.e. more personality-focused), or knowledge (*g*-focused), since these wording changes can have an impact on the degree of predictive validity, and sub-group differences (ethnicity, gender etc.)

3.68 Practical constraints may limit the use of SJTs by driving instructors in the car, but it may be possible to have slightly less formal ways of testing for

situational judgement, based on some form of in-car commentary or delayed commentary driving (see Section 4.) It is likely that formal, scored situational judgement testing will need to be carried out in the theory test, rather than the practical test, due to time constraints, and a need to have control over the situations people are asked to judge.

3.69 A further issue to be addressed is the training of instructors and examiners. They are likely to require at least basic training in the administration of SJTs (or less formal equivalents), and probably some training in the 'coaching skills' that may be required in order to coax learner drivers to examine their attitudes.

3.70 With respect to hazard perception and situational awareness (SA), it is argued here that these two can be thought of as equivalent in the driving domain. Hazard perception testing is already implemented in GB driver training and assessment, and there seems little to be gained from developing tests of SA beyond the video tests already used for hazard perception testing (performance on which has been shown to be related to accident risk).

# 4 Commentary driving

**Section 4 findings**

- Commentary concurrent with the driving task does interfere with driving performance and hazard perception, even in more experienced drivers such as advanced police drivers.
- Even if commentary does not interfere with driving, this could be because the driving task is prioritised, and commentary is neglected (therefore undermining its usefulness).

**Section 4 recommendations**

- Commentary driving (verbalising while actually driving) is not suitable for use with learner drivers.
- The use of 'delayed' commentary or questioning by examiners after a traffic event has promise, and avoids interference with driving.
- With careful construction of questions for specific traffic events, delayed commentary could serve as an informal test of situational judgement, including attitudes and motivations as well as aspects of hazard perception.
- We suggest that whether delayed commentary can be used for assessment as well as training will depend on overcoming concerns regarding standardisation, and ensuring that the knowledge being probed is the kind of knowledge that can be verbalised (as opposed to procedural knowledge, which is difficult to put into words).

4.1 In this section, three key areas of evidence relating to the suitability of introducing commentary driving into GB driver training and assessment are discussed. Firstly the extent to which learner drivers (and indeed all drivers) are able to 'multi-task' when driving and commentating at the same time, without their driving suffering, is considered. The answer to this issue would be crucial to the implementation of commentary driving in driver training, since distraction is a key contributory factor in road accidents. The second issue considered is the possibility that even if learners can be shown to cope with commentary in a safe way, they may only be able to do so if they pay full attention to the driving task (while only paying lip service to describing the contents of their awareness). If this is the case, the introduction of commentary driving would again be called into question, since it would be of little practical value. Finally, the possibility that some of the knowledge drivers possess about hazards on the road ahead is not available to conscious reflection is discussed. So called 'implicit' knowledge that is relatively difficult to verbalise is well-documented in the psychological literature, and this may have implications for the use of commentary driving in GB driver training and assessment. This final point is just as relevant to the use of 'delayed-commentary driving' techniques, as defined in paragraphs 2.17–2.21. Attempts at implementing commentary-type techniques in other countries is

discussed, and evidence from GB police driver training, in which commentary driving is used extensively, is also reviewed.

## Dual-task interference—can novice drivers be expected to verbalise while driving without making their driving suffer?

4.2 A key consideration related to commentary driving, especially for learner drivers, is how it might impact on driver workload, and thereby may interfere with the driving task. The scale of the problem of driver distraction and overload is illustrated in a paper from Underwood (2007). Underwood reports that estimates of the proportion of accidents in which distraction is a factor range from 15%–40% from accident descriptions, but that the figure is even higher if these accidents are observed directly. For example, Klauer, Dingus, Neale, Sudweeks, and Ramsey, (2006) ran the '100-car naturalistic driving study' in Virginia. In this study, 100 cars were instrumented with various sensors and video cameras, and 241 drivers used the cars over a one-year period, resulting in more than two million miles of recorded driving. During this time, the data showed that of the 82 crashes and 741 near-crashes recorded, 78% and 65% respectively involved distraction. Clearly, any introduction of commentary driving needs to consider carefully the impacts on driver workload and distraction that may result.

4.3 It is outside of the scope of this review to discuss in detail the various psychological models of workload and interference in cognitive processing that exist. Suffice to say, there is general consensus in the literature that different tasks will interfere with each other to different degrees, depending on the extent to which they overlap in terms of the types of mental processing required to perform them. Two dominant models that have been used to describe the fine details of such dual-task interference are the multiple resource model (Wickens, 1984) and the working memory model of Baddeley and Hitch (1974). The interested reader is referred to Wickens (1984) and Andrade (2001) for overviews of these models.

4.4 The important point for the purposes of the current review is that such models typically posit the existence of different sub-components of the human cognitive system (often related conceptually in neuropsychological approaches to different parts of the brain), all of which deal with different types of processing. Support for this general notion comes from the dual-task paradigm. In the dual-task paradigm, an individual is required to perform two tasks simultaneously, and performance is compared with that of the tasks in isolation. When performance on one and/or both tasks is lower when the tasks are done simultaneously than when they are done separately, these two tasks are said to interfere with each other, and it is assumed that both tasks compete for the same class of information processing resources in the brain.

4.5 A brief description of the multiple resource model Wickens (1984) helps to illustrate this general point. The model posits the existence of several different pools of mental resources that can be tapped simultaneously (see Figure 2). The degree of decrement in primary task performance induced by a secondary task that is performed concurrently varies depending on the nature of the characteristics and mental resource requirements of the two tasks. Tasks that require quite different mental resources may, according to the model, be performed simultaneously without much problem, whereas those drawing on similar resources may interfere considerably with each other.

4.6 It should be noted that the types of processing resources are not limited to input modalities. For example it can be seen in Figure 2 that as well as tasks differing

in whether the stimulus input is, for example, auditory or visual, they also differ in the stages of processing required (perception, processing, and action selection) and the level of reasoning required (subconscious, symbolic, and linguistic). As pointed out by Horrey and Wickens (2003), among others, interference between tasks is derived from overlap at any or all of these levels. This explains why two tasks related to driving that appear at face value to be very different (hazard perception, and talking on a mobile phone, which require predominantly visual and auditory inputs respectively) still interfere with each other, since they both require some elements of cognitive processing and action selection (see later in this section).

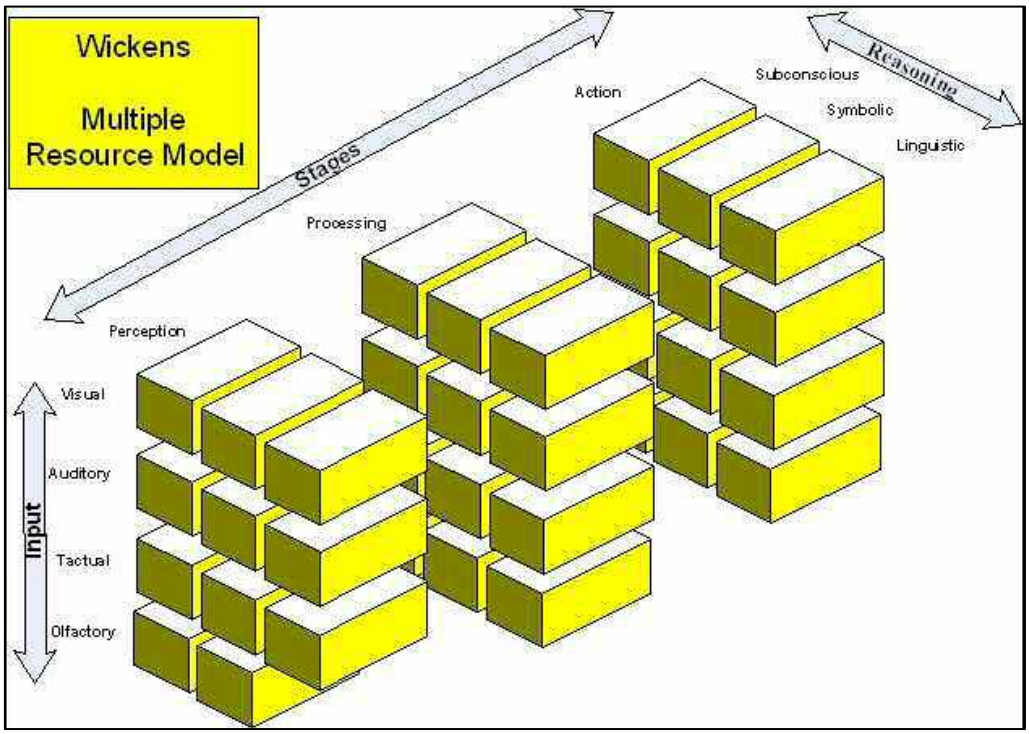


**Figure 2: Wickens' (1984) Multiple Resource Model.**

4.7 Another issue relevant here is the extent to which driving can be interpreted as a highly complex, perceptual-motor skill (Shinar, 1978). According to psychological theories of skill acquisition, perceptual-motor skills are more vulnerable to intrusion in early stages when significant cognitive control over the task needs to be exerted (Fits and Posner, 1967). With increasing practice, performance becomes automated and requires less cognitive control. This would suggest that later stages in the learning to drive process may be more appropriate for inclusion of commentary driving than earlier ones, at least in terms of the degree to which the motor-control elements of the driving task are interfered with. There is general acceptance in the literature that novice drivers do take time to develop the basic motor-control skills needed to control the car. There is also evidence from various studies. For example, Underwood, Dobson, Chapman and Crundall (2001, cited in Underwood, Crundall and Chapman, 2002) showed that differences in horizontal scanning on rural roads and dual-carriageways between novices and experts in real driving disappear when the two groups are simply shown video clips of the two types of road (thus removing the vehicle control load for novices).

4.8 A large literature now exists on driving and dual-task interference from various secondary tasks. Specifically, there are a number of studies suggesting that any kind of concurrent verbalisation seems to interfere with various aspects of driving performance, although it is noteworthy that the literature that has concentrated

specifically on the impact of commentary driving (as we have defined it here) is relatively small.

4.9 In paragraphs 4.10–4.16 evidence from dual-task interference in driving in laboratory tasks, and from evidence related to mobile phone use, is discussed. Then in paragraphs 4.17–4.22, attention is turned to further anecdotal evidence from various sources in the literature on implementations of commentary driving in driver training and in research applied specifically to learner drivers.

### *Laboratory studies of dual task interference in driving*

4.10 Important evidence for interference comes from the literature on hazard perception. This is particularly important, since hazard perception is the one skill that has been shown, across a number of studies, to be related to accident risk (see previous sections). There are at least two studies that have shown a decrement specifically in video hazard perception skill when secondary tasks are used. McKenna and Farrand (1999) had drivers carry out a random number generation task while looking for hazards in McKenna's video hazard perception test. It was found that hazard perception scores were around 10% worse in the dual task condition, compared to the single task (i.e. hazard perception only) condition for novice drivers, and an even larger decrement (approximately 25%) was observed for more experienced drivers. Rowe (1997), in unpublished work, found similar results using letter detection as the secondary task.

4.11 The finding in the McKenna and Farrand study, that experienced drivers (whose hazard perception scores were higher than those of novice drivers in the single task condition) suffered even more when carrying out the secondary task, was interpreted by Horswill and McKenna (2004) as compatible with the interpretation of hazard perception as being an effortful skill that continues to develop, but can only convey a road safety advantage if sufficient attentional resources are allocated to its ongoing use. This point is returned to in paragraphs 4.49–4.51 when we discuss the use of commentary driving in advanced GB Police driver training.

4.12 A number of other studies suggest that various types of verbal secondary task interfere with various aspects of driving performance. For example, Kass, Cole and Stanny (2007) showed that a simulated hands-free phone conversation impacted on SA for driving as measured by SAGAT-type SA measure in a simulator. Lee, Lee and Boyle (2007) had drivers carry out a dual task of responding verbally to questions regarding previously-heard information while driving in a simulator. They found that this task interfered with the attention paid to pedestrians in potentially hazardous positions on the road. Importantly however this study showed that the task only interfered with attention to the pedestrians when drivers were actually responding to the questions, rather than simply listening. This is important for two reasons. Firstly, it further supports the prediction from the multiple resource model (Wickens, 1984) that tasks sharing a response or 'action' component will interfere with each other even if the input modalities are different. Secondly, it is highly relevant to commentary driving, which by its definition is an active process.

4.13 Another relevant study comes from Redelmeier and Tibshirani (1997), who studied accident-involved drivers over a 14-month period, and examined their mobile-phone call history. They found that the use of a mobile phone (hands-free or hand-held) increased the risk of an accident four-fold. Other studies into the use of mobile phones (even hands-free) have shown similar increases in accident risk (e.g. Burns, Parkes, Burton, Smith and Burch, 2002).

4.14 What about research using actual commentary driving as the secondary task? One study is that of Crundall and Underwood (1997). The authors studied the effects of commentary driving directly in terms of its effect on visual scanning patterns, and on hazard perception. Subjects did either 'natural reports' or 'restricted reports' which were only one or two words and restricted entirely to things that were being attended. Participants all had driving licenses. No differences found between conditions on hazard perception performance, or in terms of mean fixation duration for the whole video footage.

4.15 On first inspection, these results would seem to run counter to the results obtained using other types of verbal secondary tasks as mentioned above. However, one possible explanation for the Crundall and Underwood (1997) results is that the hazard perception test used was not sensitive enough to pick up the differences in reaction time that might be expected. It is usual in hazard perception studies to report validation data (such as a test's ability to discriminate between novice and experienced drivers—e.g. McKenna and Crick, 1991; McKenna and Horswill, 1999), but no-such data were reported in this study. It should also be noted that even if the commentary did not interfere in the laboratory (implying it is a less disruptive task than some of the other tasks used in McKenna's laboratory, such as random number generation), this does not mean that it would not interfere on the road, when drivers were having to devote some attentional resources to controlling the vehicle.

4.16 Thus, the literature reviewed in this section shows that secondary tasks generally do interfere with hazard perception and other aspects of driving performance. However, the tasks used are usually related to something other than the driving task itself. Although Crundall and Underwood (1997) demonstrate that commentary driving itself did not interfere with eye movements and hazard perception in their study, more research is needed to overcome the methodological problems with this study, and to quantify further the possible interference effects of commentary driving.

### *Operational implementation of commentary driving—general issues*

4.17 In addition to the evidence reviewed in paragraphs 4.10–4.16, we can also consider what practitioners and applied researchers have suggested regarding the implementation of commentary driving. This more 'anecdotal' evidence from domain experts is likely to prove useful in any final consideration of the use of the technique in driver training and assessment.

4.18 In general it would appear that the concern regarding commentary driving potentially interfering with driving is reflected in the operational implementation of such techniques. For example, Baughan et al. (2005) interviewed DSA test examiners regarding the inclusion of commentary techniques, and although the potential value of the technique was recognised, several problems were identified, including:

- candidates would find it difficult to use the technique as it is difficult to learn;
- candidates would find it difficult to commentate on hazards as they find it difficult to perceive hazards; and
- examiners would still have to keep an eye on basic driving controls even when listening to commentary.

4.19 Respondents felt that stopping the car to allow retrospective commentary may solve some of these issues, but that there was no time in the test to achieve this (and see paragraphs 4.36–4.47 for problems with this approach). McCormack

(2003) reported a training programme for hazard perception for learner drivers run by a driving school in the UK. As part of the training, an on-road commentary was proposed, but it was acknowledged that in the very early stages of driving, when learners may not have automated some of the basic control procedures of the car (such as changing gear), instructors themselves should drive and comment, inviting learners to observe and comment themselves without being in control of the car.

4.20 Gregersen (1994) evaluated the impact of three different teaching techniques on accidents over two years in a sample of learner drivers. The training included commentary driving as one element. The effect on accident risk was negative in the first year; drivers who underwent the training had more accidents than the control group. In the second year, they had fewer. Gregersen suggested that the high cognitive load experienced by new drivers while they are still grappling with control skills may prevent them from using the training techniques until they have automated enough control tasks (gear shifting etc.) to free up the resources they need.

4.21 Wilde (1993) also suggests that commentary driving may interfere with control for beginner drivers, but suggests that it should be simplified for beginners and focused more on risk. Wilde describes a technique where drivers simply report a rating on a 1–7 scale when they perceive a change in risk when driving. Again, the focus is on reducing the chance that the technique will interfere with driving performance.

4.22 Evans and Macdonald (2002) are also mindful of the possible interference effects from commentary to driving. Their research was designed to document the mental 'schema' of pre-learners and probationary learners of differing levels of skill. These drivers were driven around a course in Adelaide, either at night or during the day. This setting ensured that the participants did not need to commentate and drive themselves at the same time. At night, certain objects such as reflective items attracted attention and this showed up in the commentaries, even for the pre-learners. The finding suggests that the commentary technique, even when carried out by people who are not actually driving, can yield interesting results regarding which task-relevant features of the environment drivers are attending to.

### *Operational advice on commentary driving in Australia*

4.23 Not all implementations of commentary driving in training are as cautious of compromising safety as the ones reviewed above. For example, in the Handbook of Hazard Perception (Roads and Traffic Authority, NSW, Australia, 2001) drivers who have attained their provisional licences are advised to practise concurrent verbalisation while driving and to compare their verbalisations with those of a more experienced driver, in order to help improve their hazard perception skills. Although the handbook suggests that drivers pull over to a safe place to discuss verbalisations with their more experienced colleagues, the act of verbalising is clearly intended to occur while the novice driver is in control of the vehicle, and no mention is made of taking care of cognitive overload.

### *Operational implementation of commentary driving in the Netherlands*

4.24 Recently, Vissers, Mesken, Roelofs and Claesen (2008) have described pilot work to add situational judgement-type testing (among other techniques such as independent driving) to the Netherlands' practical driving test. Questioning by

the examiner and commentary driving were both piloted. The 'questioning' of candidates regarding problematic traffic situations (delayed commentary) was well-received by candidates and examiners, although it was noted that any candidates with language problems may find it difficult to offer suitable answers, and also that examiners may interpret answers inconsistently, thus threatening the reliability of such methods. Commentary driving was judged by all those involved as very difficult for learners to perform. Participants also agreed that it certainly was not feasible as an assessment instrument, although they suggested that it may be useful as a training instrument.

***Summary***

4.25 On the whole, there seems to be acceptance in the practice of commentary driving that the technique will inevitably have some deleterious effects on driving performance in (learner) drivers. It is also worth noting that no implementation of the technique has attempted, explicitly, to cover higher levels of the GDE matrix (such as 'skills for life'), or the notion of 'calibration' (see paragraphs 3.25–3.38). It could be argued that this technique has potential as a 'stripped-down' version of full SJTs that can be used at the roadside, and conceptually there is no reason why the probes for such verbalisation cannot focus on knowledge from the higher levels of the GDE matrix. This point is returned to in the recommendations section.

## Is commentary driving a good indicator of the mental processes underlying driving performance?

4.26 The discussion of commentary driving has, so far, focused on the issue of dual-task interference—the possible deleterious effect of commentary on driving performance. A number of studies have been reviewed showing that, especially in the case of hazard perception, concurrent verbal tasks interfere with driving skill and ability. This is the key safety issue relating to the possible introduction of commentary driving into GB driver training and assessment. If it is done at all, it will need to be introduced in such a way as to minimise the impact on safety, and driving instructors will need to be wary of the increased danger when their students are engaged in commentary.

4.27 However, there are two additional considerations directly relevant to the suitability of commentary driving—both linked to how accurately the contents of a driver's awareness can be reflected in their commentary, even if a version of commentary driving can be found that does not interfere with the driving task.

4.28 Firstly, there is the flip-side of the dual-task interference problem—the possible deleterious effect of driving on the commentary task itself. It is possible that learner drivers may simply prioritise the driving task at the expense of the commentary task, thus meaning that commentary may not provide as accurate an insight into the contents of their awareness as we would like. Secondly, there is the issue of whether commentary itself can ever be an accurate reflection of all the contents of a driver's awareness, given the difficulty people have in verbalising all of their thought processes. Both of these issues are considered below.

### *Interference from driving to commentary*

4.29 Even if the driving task did not suffer when concurrent verbalisation is taking place, concurrent verbalisation performance might suffer as a result of drivers prioritising the driving task, meaning that it may not be as reliable an index of the contents of awareness as one may hope. This problem of trade-offs between primary and secondary tasks is that a major problem in research using dual-task paradigms to investigate interference effects.

4.30 There are several examples in the literature that relate directly to the specific issue of concurrent verbalisation when engaged in driving tasks. For example, Bowdich and Groeger (2002) showed that although a concurrent verbal reasoning task did not interfere with performance on a real driving task (acceptance or rejection of gaps between traffic cones on a test track), the driving task itself interfered with concurrent performance on two tasks (verbal reasoning, and paced auditory serial addition) known to measure central cognitive functioning. When these tasks were carried out during a drive, accuracy was between 7% and 11% lower than when they were carried out alone, while the driving task itself (driving between cones) did not suffer. The results are consistent with the interpretation that drivers were prioritising the driving task over the concurrent verbal tasks.

4.31 Another relevant study comes from Spence and Read (2003). Participants in a driving simulator were asked to 'shadow' (i.e. repeat) an auditory message coming either from speakers to their side, or from their front, while driving. Although Spence and read found no interference from the concurrent verbalisation to the driving measures used (measures such as lane position, speed, mean decision time merging at junctions, and gap acceptance), they did find that the driving task itself interfered with shadowing—especially when the message being shadowed came from the side (30% drop in accuracy) rather than in front (11% drop in accuracy) of the driver. The difference in interference between the two directions of shadowing is consistent with the interpretation that auditory attention and visual attention are typically focused in the same direction.

4.32 One problem with studies such as Bowdich and Groeger (2002) and Spence and Read (2003) for current purposes is that they use tasks that are similar conceptually to commentary driving, but which are not exactly the same. For example, in both the studies mentioned above, the focus of the secondary verbal task is something different from the driving itself. It is possible that commentary driving, with a focus on the visual stimuli to which the driver needs to be attending anyway to perform the driving task, will not experience the same level of interference from driving. There is tentative evidence in Spence and Read (2003) for this possibility—driving interfered less with shadowing a message from the front than with shadowing one from the side. Although commentary driving (unlike shadowing) does not require auditory attention, this finding does suggest that when the aims of the driving task and the verbal task are more 'aligned' there seems to be less interference.

4.33 There are at least two studies providing data that are directly relevant to the issue of interference from the driving task on the performance of commentary driving. Smith and Murdoch (1986) studied the use of a commentary driving technique that is applied to the safety evaluation of low volume rural (LVR) roads in the US, with a specific focus on identifying information-deficient sections of the road. Thus the technique is not used to identify immediate hazards to the driver, but the information content of different sections of the road. Students being taught the technique carried out commentary while they drove the route, or while they watched video tape of the route. Their identification of problems was scored

by an experimenter (an expert who had previously scored the routes). Average scores were higher in the video group compared to the driving group. Since the commentary was of the kind that might be expected to be performed by drivers undergoing tuition (i.e. driving related), this finding can be taken as direct evidence of the driving task itself making performance on commentary driving worse.

4.34 Another relevant study comes from Hughes and Cole (1986), who had participants verbalise what attracted their attention while driving a route or watching film of the same route. They showed that the level of verbalisation for 'traffic control devices' and also for 'vehicles' was higher in the video condition (nearly 40% more verbalisation for traffic control devices, and 76% more for vehicles, in the video condition). Again this is consistent with the interpretation that the driving task interfered with the ability of drivers to use commentary, possibly because the driving-related processes took precedence and prevented verbalisation on some occasions. Interestingly, the version of commentary used in the study was fairly simplistic, in that drivers were only asked to mention things that grabbed their attention—they were not asked to offer detailed descriptions of hazardous situations or how they were going to deal with such situations. Hughes and Cole (1986) advocate this version of commentary driving as being more suitable, especially for learner drivers, although their own data are compatible with the interpretation that even for this simple version of the technique, interference from the driving task can occur.

4.35 Overall, the evidence supports the idea that even in those circumstances when commentary does not make driving worse, this may only be because drivers ensure that sufficient prioritisation is given to the driving task, at the expense of the commentary. There is reason to believe that this effect would be considerably stronger in learner drivers, who will still need to pay more attention to the basic driving tasks than more experienced drivers typically used in the research cited in this section. Therefore, the possibility that learner drivers may not be able to use commentary effectively needs to be considered.

### *General problems with introspective methods of assessing awareness*

4.36 An important question regarding the effectiveness of commentary driving is whether the knowledge underlying driving performance is available to conscious, verbal reflection. A well-established finding in the cognitive psychology literature is that often the knowledge underlying tasks—especially complex tasks—is not fully available to conscious reflection, despite it being possible to show that the knowledge is indeed influencing performance on a given task. The theoretical constructs involved in such discussions are many, but include: procedural versus declarative knowledge (e.g. Baddeley, 1990); implicit memory (e.g. Roediger and McDermott, 1993); implicit learning (e.g. Berry, 1997); and perception without awareness (e.g. Merickle, 1992).

4.37 It is outside of the scope of the current review to go into a detailed discussion of the questions that have been debated in these literatures. However, there are two key issues that have direct relevance to the way in which any kind of commentary driving technique—and for that matter, any techniques designed to access the contents of awareness—is introduced, especially if learner drivers are actually assessed on these measures.

4.38 A first point is the basic finding that people are actually rather bad at reporting accurately on the stimuli that influence their behaviour, even in cases when the tasks are not 'procedural' which would not be expected to be amenable to

conscious reflection (e.g. Baddeley, 1990) anyway. In a seminal paper, Nisbett and Wilson (1977) reviewed evidence from the literature, and showed that generally, when people are asked to verbalise the stimuli that affect their responding across a number of different tasks, they do so not on the basis of a true introspective awareness of all the relevant stimuli and responses underlying their behaviour, but actually on the basis of *a priori* causal theories about which stimuli are plausible causes of their behaviour (and only when these stimuli are salient enough to enter consciousness). Often, people are unaware of the stimuli that cause them to show specific behaviours, or are unaware of the behaviour, or are unaware that a particular stimulus has caused a specific behaviour.

4.39 An example in driving might be as follows: a driver giving a verbal commentary as he drives may fail to comment on an oncoming left-hand bend in the road, but behaviour (moving towards the centre of the road so as to get a better view of the road ahead) would indicate sensitivity to the presence of the bend. When asked later to describe the reason for moving toward the centre of the road, failing to recall the bend the driver may state that he did so to put himself further away from some parked vehicles, or from some pedestrians (both plausible reasons for the behaviours, but maybe not the real reason at the time).

4.40 This basic finding, in a number of forms, has been shown in multiple settings across a number of literatures, as listed above. This mismatch between what is known, and what can be reported, has also been examined in driving tasks. For example, in Gugerty's study (1997), participants watched animated driving scenes showing both forward and rear views from 'their' vehicle. Participants had to avoid some vehicles driving in a hazardous way through simple manoeuvring of their own vehicle. Gugerty found moderate correlations between explicit recall measures of SA (for car positions) and implicit measures (manoeuvring), suggesting that at least some of the processing underlying manoeuvring may be unavailable to awareness.

4.41 The degree of mismatch between performance and awareness is a contentious issue in the literature, and this will not be discussed in detail here. The important conclusion is that there is often a mismatch between performance and awareness, and this has implications for testing of awareness in complex tasks like driving.

4.42 The second, more pragmatic issue is again related to the mismatch between performance and awareness. In a key review paper, Shanks and St John (1994)[14] suggested that rather than entertaining the possibility that people can use information to guide their performance, but remain unaware of it, all dissociations between performance and awareness could be explained by two criteria on which the measures used for each (e.g. for our purposes here—driving, and verbal report of driving) differed. These two criteria were termed the 'sensitivity' criterion and the 'information' criterion'.

4.43 The sensitivity criterion states that when performance and awareness tests show dissociation (e.g. knowledge is recalled through a performance, but not through something like verbal report), this does not mean that people are not aware of the knowledge they are using to guide their performance. Rather, the measures of performance and awareness might simply differ in their ability (or sensitivity) to access this knowledge. An example from driving will help to illustrate this.

[14] This reference goes into a great deal of theoretical detail about the differences between 'implicit' and 'explicit' learning. This is outside of the scope of this review. However, the central point made in Shanks and St John (1994)—regarding the dangers of inferring too much from dissociations between 'performance' and 'awareness'—is key to the proper assessment of learner drivers' awareness of road hazards through techniques such as commentary driving, and even situational judgement.

Imagine that a learner on their test is driving towards a two-lane roundabout, and has been asked by their examiner to drive straight on. Ordinarily, the left hand lane should be used for this manoeuvre. However, let it be assumed that a following driver is driving very close to the learner, and shows intention to overtake the learner on the left at the entrance to the roundabout. Because of memory decay, asking a learner driver to verbally report their memory for this traffic event ("Please tell me what you remember about the hazards and risks in the roundabout we drove through a few minutes ago") might not reveal any apparent memory for the following driver. On the other hand, the learner's performance in the situation at the time might reveal that knowledge of the following driver is indeed guiding performance (e.g. positioning their car slightly to the inside lane to avoid the following driver). Thus if an examiner wished to check why the learner behaved is a certain way (in this example, why they positioned their car 'incorrectly' on the right-hand lane of the roundabout, when they should have been on the left), the examiner might come to the conclusion that the learner simply made a mistake, rather than coming to the correct conclusion that the learner saw the following driver, and performed appropriately to avoid a conflict, while being unable to verbalise this fact due to decay in memory.

4.44 The second criterion suggested by Shanks and St John is the 'information' criterion. This states simply that when a dissociation is obtained between a measure of performance and a measure of awareness, it might be due to the fact that the measures are asking about different things. Again an example from driving will illustrate how this issue could lead to inaccurate assessments of what learner drivers are aware of (or not). Imagine that a learner is approaching a pedestrian crossing, that the driver's view is largely blocked by a parked lorry, but that the driver catches a glimpse of a pedestrian's legs underneath the lorry, indicating that the pedestrian is beginning to cross. The driver may slow down in this situation. Later on, if the instructor wishes to gauge the learner's awareness of the situation, he may ask "What was it about the lorry that caused you to slow down back there", and may receive the answer "Nothing". The learner, of course, is actually thinking that it was the pedestrian that caused him to slow down, but without further questioning, the instructor may erroneously conclude that the learner was unaware of the hazards present.

4.45 This issue has also been raised by Chapman and Underwood (1997), and these authors also present data (admittedly over longer time periods) demonstrating the ways in which recall of accidents and near-accidents can decay differentially over a two-week period. The least serious near-misses (self rated) were the ones most likely to be forgotten, which has implications for what can be expected from retrospective commentary techniques for relatively innocuous driving events during a lesson or test. They also point out that some types of memory decay more quickly than others (e.g. Mohr, Engelkamp and Zimmer, 1989; cited in Chapman and Underwood, 1997).

4.46 Another finding relevant to the issue of retrospective memory for driving events comes from Chapman and Groeger (1992), who showed that subjective risk distorts memory for traffic situations (recall on the road and recognition in the lab). Note that there is no explicit description of delays between driving/observing video clips and recall/recognition in this paper, but it is estimated that in both cases the delay would be around 15–30 minutes from the middle of the drive/viewing to the surprise recall/recognition test. Evidence was found for the 'inverted U' hypothesis (Deffenbacher, 1983), whereby extremely un-arousing, and extremely arousing traffic events were recalled less well than those traffic events that were moderately arousing. Evidence was also found supporting the Easterbrook hypothesis (Easterbrook, 1959), whereby the focus on 'central'

details of traffic events increased in memories for those events, and the focus on 'peripheral' details decreased, as arousal of those traffic events went up. This has implications for drivers learning from traffic situations—especially at delays of those likely to have been present in this paper. The situations that examiners and instructors might most want to tap into in terms of what learner drivers were aware of may be the ones (i.e. the arousing ones) that drivers have least recall of.

4.47 Much the theoretical work being reviewed in this section is rather 'academic'. However, it reveals an important consideration for the measurement of the contents of awareness in learner drivers (or any drivers for that matter), whether by concurrent commentary driving, or by retrospective questioning (i.e. delayed commentary). When there is an apparent mismatch between what drivers do, and what they can verbalise about why they did it, it may not be the case that they were erroneously unaware of the relevant parts of the road environment. Instead, it may simply be the case that a driver is unable to verbalise some or all of the knowledge that guided his behaviour, either because it is not the kind of knowledge that can be verbalised, or because the measures being used to probe awareness differ from the performance measures in terms of their sensitivity, or the information they are asking about. Careful consideration of these issues will need to be included in any development of such measures of awareness in future GB driver training and assessment.

## Applicability of commentary driving to different demographics and personality types

4.48 Commentary driving requires certain standards of language production skills, and an ability to express oneself without feeling self-conscious. In fact, feedback from police driving instructors (see paragraphs 4.49–4.51) suggests that these are some of the main pre-requisites for police drivers who are able to master the skill. If commentary driving is to be introduced into GB driving assessment, it will be necessary to ensure that particular demographic groups (e.g. those with poor language skills, or high levels of shyness) are not unfairly disadvantaged.

## Evidence from police advanced driver training

4.49 Commentary driving is used extensively in GB police driver training. Police driver training consists of three different 'levels':

1. Basic: this covers being simply able and allowed to drive a police car at an acceptable level, slightly above DSA standard.
2. Standard: 4-week course covering more work based skills such as driving a police car on an emergency response using blue lights.
3. Advanced: 4-week course covering specialist skills such as high-speed pursuit, and also firearms quick response. Advanced drivers should be able to plan their drive to a much higher standard.

4.50 Commentary driving is not covered at all at the basic level of training. It is introduced in second week of the standard level, and developed throughout the rest of the standard and advanced courses.

4.51 Interviews with Police driving instructors and questionnaires given to some Police driver trainees were used to gather evidence from these expert groups pertaining to the use of the technique in learner driver training and assessment.

### *Interviews with police driver instructors*

4.52 In February 2008, two semi-structured interviews were carried out with police driving instructors, both of whom have experience teaching at the above levels. In these interviews, views on commentary driving, including its potential use in learner driver training and assessment, were elicited. The specific questions asked are listed in Appendix C.

4.53 A summary of the results from these interviews is presented in the bullet-points below:

- The focus of the skill is picking out hazards, prioritising them, and planning how to deal with them. Additionally, trainees are told that the skill is required to allow the instructor to understand what the trainees are thinking about when they are driving.
- When the skill is first taught, it is introduced in a hierarchical fashion. For example, first drivers are taught to commentate on the 'start-up' procedure of the car, and on general features such as the weather conditions. Later, drivers are encouraged to comment on the direction the road is taking; then on identification of junctions; and so on. Sometimes the drivers are asked questions about the environment to prompt them to particular parts of that environment. This 'easing into' commentary is done to help develop the skill of multi-tasking the specific behaviours of 'commentating' and 'driving' before full commentary commences.
- In some forces, commentary is very structured—specifically around the standard 'system of car control' used by police drivers, which concentrates on: information; position; speed; gear; and acceleration. Also, some forces use commentary to encourage trainees to scan their environment in the order 'far, medium, near', thus encouraging early sight of potential hazards.
- Driving behaviour is definitely altered by commentary driving, especially in drivers when they first begin using the technique, but sometimes even in advanced drivers. Examples of alterations in behaviour include reduction in speed, stalling, selecting the wrong gear, excessive hesitation, and 'overlapping' driving actions like steering and changing gear (i.e. doing them at the same time—something that Police drivers are taught not to do to improve safety). However safety is not compromised due to the fact that the police drivers are all sufficiently skilled and experienced in the basic controls of the car, and are under the supervision of an experienced trainer.
- Given that even advanced police drivers change their behaviour when engaged in commentary, it was suggested that commentary driving would be extremely difficult for learner drivers to perform without considerable changes to, and possibly decrements in, their driving performance.
- Police drivers on the Standard course experience approximately 40 minutes of training in the technique per day, every day for two weeks. Because the drivers are trained in threes (taking it in turns behind the wheel), they also have around twice this time to observe other drivers practising the technique. This experience is sufficient to permit most drivers to master the technique.
- 'Ability to multi-task' and 'not being self-conscious and shy' are two attributes thought to be the key pre-requisites needed to master the technique.
- It was felt that, although in principle it is a laudable idea, to introduce commentary driving into learner driver training and assessment would require a much greater amount of time to be built into the training protocol.
- It was suggested that a more practical way of gaining some of the benefits of commentary driving would be to either rely on the 'question and answer'

approach, or just to show learner drivers how commentary could help to structure their own observations (e.g. the 'far, medium, near' approach).

- It was also felt that learners would have to be near to test standard to benefit from the technique, since it is only when basic control skills have been largely automated that they would be able to free up the attention needed to multi-task. It was felt that introducing the technique too early in driver training would result in either a decrement in driving safety, or an inability to commentate properly due to having to divert all attention to the driving task.

### *Questionnaires given to police driver trainees*

4.54 A short questionnaire (see Appendix D) was distributed to several police driver trainees at a police driver training school. The following bullet points summarise the main findings (11 responses were obtained):

- On average, respondents rated the difficulty of the technique as 3.9 out of 5 when they first started learning it, but 2.1 out of 5 after some experience with the technique.
- Most respondents stated that they were apprehensive about the technique before they did it, and that they expected it to be challenging.
- Most respondents stated that they thought it was a good training tool for 'widening awareness of the road'.
- Only one respondent mentioned that their driving actually suffered as a result of performing commentary. The other respondents generally stated that it actually helped in terms of their 'awareness' and 'observation'.
- Most respondents mentioned that the most difficult thing about commentary driving is prioritising what to include in commentary, and what to leave out.

4.55 In general, the responses of the police trainee drivers back up what the police instructors said about the technique of commentary driving.

## Summary

4.56 The evidence reviewed in Section 4 seems to suggest that although commentary driving has some potential for use in GB driver training, there are likely to be some serious issues related to its interference with the driving task. It should certainly not be used early in training as learner drivers will not be able to free-up the required mental resources while they are still learning 'the basics'. Thus either the commentary driving will interfere with basic perceptual-motor control aspects of the driving task, or drivers will prioritise the driving task, and will not be able to give appropriate commentary.

4.57 Additionally, even later on in training, it should be accepted that even with very basic 'bullet point' techniques (i.e. giving quick keywords to describe those aspects of the road environment that are drawing attention), some changes in driving may become apparent; especially decrements in hazard perception. Another option is to have learners use commentary when they are not in control of the vehicle (delayed-commentary driving), for example while the car is parked safely after the traffic event of interest has been dealt with. The technique clearly has potential, especially in this latter situation, to be used as an informal version of an SJT.

4.58 Whatever the level of implementation, the technique would appear to have most use in allowing driving instructors to gain an insight into those aspects of the road environment of which learner drivers are aware. This information can be used to help with feedback to learner drivers with regard to their observational skills. Care needs to be taken, however, to avoid the assumption that a learner did not notice a particular feature of the road environment just because he did not mention it. It may in fact be the case that verbal report is not sensitive enough to pick up some knowledge of which a driver may have been unaware (or only dimly aware), but which nonetheless influenced driving appropriately.

4.59 Commentary driving is unlikely to be useful during assessment for leaner drivers, except as a way for the examiner to gain similar insights into a learner driver's awareness so that examiners may pass on advice, increasing the extent to which candidates get the chance to 'learn' from their driving test. The difficulties associated with setting a pass/fail criterion for the technique (i.e. difficulty of arranging standardised scenarios; possible mismatches between contents of commentary and knowledge being used to maintain performance) probably preclude it from being used a formally-assessed part of the test (not even police driving instructors formally assess the skill in this way).

# 5 Overall conclusions and recommendations

5.1 This review has focused on evidence relating to the suitability of situational judgement testing, and commentary driving, for use in GB driver training and assessment. Evidence from the relevant literatures of employment selection, on the related concepts of situational awareness and hazard perception, and general cognitive psychology, has been considered. Applied evidence from EU projects and from other countries has also been reviewed, where attempts have been made to introduce elements such as situational judgement testing and commentary driving into driver training and assessment. Finally, specifically related to commentary driving, Police driving instructors and trainees have been interviewed, to assess practical implications of that technique. The overall conclusions, and recommendations, are listed below.

### Findings

The key findings in the review are as follows:

5.2 Situational judgement is usually measured through the administration of formal Situational Judgement Tests—usually in the format of a description of a situation, with alternative behaviours (varying in appropriateness as rated by experts) for respondents to choose from.

5.3 Such tests require considerable effort to create and administer correctly, in order that the data obtained are valid predictors of performance in the domain in question.

5.4 Situational judgement testing has been shown to have several advantages over traditional selection instruments such as interviews, and cognitive ability testing, including better prediction of job performance.

5.5 This increased predictive power is probably due to the fact that Situational Judgement Tests assess the behavioural tendencies and personality factors needed to succeed in a job, as well as the knowledge required.

5.6 Situational judgement tests appear to be culturally fairer to minority (e.g. ethic) groups, than tests of cognitive ability.

5.7 The way in which situational judgement test items are worded is especially important since it can alter what the test is measuring (knowledge, or behavioural tendencies).

5.8 Formal situational judgement tests are unlikely to be suitable for use in practical driver training and assessment. They are more likely to have use in theory testing, where control can be kept over the situations presented.

5.9 There are some examples of situational judgement-like tests being introduced in the practical driver assessment of other countries; however there is, as yet, no validated test of situational judgement for driving.

5.10 Even if situational judgement testing is not currently feasible as a pass/fail metric, there is still value in inclusion of such testing in driver training and assessment, since it draws attention to the attitudes and behavioural tendencies that new drivers should be showing in order to be safe on the road.

5.11 By drawing the attention of drivers to their attitudes and perceived skills, those who are badly calibrated (i.e. by overestimating their skills and underestimating the demands of the driving task) should improve and road safety benefits should follow.

5.12 The concepts of hazard perception and situational awareness are related to situational judgement, in that they also address drivers' mental processing and awareness of road hazards. However, we argue that situational judgement can be thought of as distinct from situational awareness and hazard perception, since situational judgement goes beyond awareness of hazardous road situations and specifically includes attitudes and behavioural tendencies regarding driving.

5.13 It is also argued that situational awareness, and hazard perception in driving, can be seen as essentially equivalent; i.e. hazard perception can be seen as situational awareness for driving situations.

5.14 Commentary driving, when carried out concurrently with the driving task itself, is difficult, and associated with a decrease in driving performance. This is true even for highly skilled (police) drivers.

5.15 Another challenge for commentary driving as a technique is that driving test candidates may not be able to perform it sufficiently well for it to be of use; candidates may prioritise the driving task at the expense of the commentary. Furthermore they may not be able to put some of their driving knowledge into words—the research suggests that people are bad at verbalising knowledge that underlies complex skills such as driving.

## Recommendations

The following recommendations are made based on the main findings from the review:

5.16 Situational judgement testing has potential for use in driver training and assessment. If worded appropriately, such testing will be especially suited to probing the behavioural tendencies in given situations as well as their knowledge of road hazards (i.e. hazard perception or situational awareness). Therefore, DSA should consider developing and evaluating situational judgement tests for driving.

5.17 If this is to be done for practical driver assessment, then the practical implications of such a technique will need to be assessed before implementation. This includes assessing the degree to which such tests (and their scoring) can be standardised within a dynamic and changing practical test environment, and the degree to which time constraints make the tests suitable for use in the time available on the practical test. Less formal methods of testing situational judgement, such as verbal descriptions of traffic situations experienced during training and on test ('delayed commentary') are likely to be the most suitable for the practical driving test, while formal situational judgement tests are mainly suitable for theory testing where there is an opportunity to completely control contents.

5.18 If it is the case that less formal methods used on practical driving tests cannot be standardised, then they should not be used for 'assessment' and 'pass/fail' decisions; rather they should be used to draw attention to the issue of 'calibration', so that learner drivers can be encouraged to match their driving behaviour to the driving-task demands, and the risks present in the driving environment.

5.19 The introduction of some kind of situational judgement testing (formal or informal, assessed or not-assessed) should be accompanied by consideration of the good-practice methods for test creation outlined in the review—especially item wording. Ignoring such issues will undermine the effectiveness and validity of such testing in predicting safe driving outcomes.

5.20 There is a wide range of evidence that verbal secondary tasks (especially those that involve production of speech) adversely impact on driving performance. Commentary driving (verbal commentary while actually driving a vehicle) is therefore likely to have deleterious effects on performance, including effects on car control and hazard perception, at least during early stages of learning, when car-control skills are still being acquired. Feedback from Police driving instructors echoed this point of view. Any development of commentary driving should therefore proceed under the premise that it should either be carried out after a particular situation, rather than during actual manoeuvring, or it should be carried out by candidates while someone else (i.e. driving instructor) drives the car.

5.21 Further work on assessing the likely benefits of such 'delayed' commentary driving (even if only as a method of measuring situational judgement or hazard perception 'on the road') is necessary, along with work to quantify the level of interference with the driving task that results from commentary driving. This work will permit decisions to be made as to the suitability and usefulness of commentary driving in the future.

# 6 Acknowledgements

The work described in this report was carried out in the Safety, Security and Investigations Group of TRL Limited. This report and the project have been reviewed according to the TRL Technical Referee procedure.

# 7 References


Andrade, J. (ed.) (2001). *Working memory in perspective*. Psychology Press. Hove.

Armsby, P., Boyle, A. J., and Wright, C. C. (1989). Methods for assessing drivers' perception of specific hazards on the road. *Accident Analysis and Prevention* 21(1), 45–60.

Baddeley, A. D. (1990). *Human memory: theory and practice*. LEA. Hove.

Baddeley, A. D., and Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation*, 47–89. New York: Academic Press.

Banbury, S., and Tremblay, S. (eds.) (2004). *A cognitive approach to situation awareness: Theory and application.* Aldershot; Ashgate Publishing, 155–175.

Baughan, C. J., Gregersen, N. P., Hendrix, M., and Keskinen, E. (2005). Towards European Standards for Testing (TEST): final report. CIECA Brussels. ISBN: 90-76408-12-12.

Baughan, C. J., and Keskinen, E. (2005) *Meeting the needs of novice drivers.* In: Baughan, C. J., Gregersen, N. P., Hendrix, M. and Keskinen, E. (2005). *Towards European standards for testing. Final report.* CIECA, Brussels. ISBN: 90-76408-12-12.

Baughan, C. J., Sexton, B., Maycock, G., Simpson, H., Chinn, L., and Quimby, A. (2005). Novice driver safety and the British practical driving test. TRL report 652. Wokingham, Berkshire: Transport Research Laboratory (TRL).

Berry, D. C. (1997). *How Implicit is Implicit Learning?* Oxford University Press.

Borkenau, P., and Ostendorf, F. (1989). Descriptive consistency and social desirability in self and peer reports. European Journal of Personality, 3, 31–45.

Bowditch, S. C., and Groeger, J. A. (2002). Driver distraction: a replication and extension of Brown, Tickner & Simmons (1969). In G. B. Grayson (Ed.), *Behavioural research in road safety 11*, 109–115. Wokingham, Berkshire: Transport Research Laboratory (TRL).

Burns, P.C., Parkes, A., Burton, S., Smith R. K., and Burch, D. (2002). How dangerous is driving with a mobile phone? Benchmarking the impairment to alcohol. TRL report 547. Wokingham, Berkshire: Transport Research Laboratory (TRL).

Chapman, P.R., and Groeger, J.A. (1992). Subjective risk and subsequent memory. In G. B. Grayson (Ed.), *Behavioural research in road safety 2*, 44–52. Wokingham, Berkshire: Transport Research Laboratory (TRL).

Chapman, P.R., and Underwood, G. (1997). Reporting and forgetting accidents and other driving events. In G.B. Grayson (Ed.), *Behavioural Research in Road Safety 7*, 120–129. Wokingham, Berkshire: Transport Research Laboratory (TRL).

CIECA (2002). *EU ADVANCED Project: Description and analysis of post-licence driver and rider training.*

CIECA (2007). *EU HERMES Project: Coaching and optimal communication skills for driving instructors.*

Cooper, C. (2002). *Individual differences.* Arnold. London.

Creighton, P., and Scott, N. (2006). An introduction to Situational Judgement Inventories. *Selection and Developmental Review*, 22(3), 3–6.

Crick, J. L., and McKenna, F. P. (1991). Hazard perception: can it be trained? *Behavioural research in road safety II.* Proceedings of a seminar, 17–18 September, Manchester University, 100–107.

Crundall, D. and Underwood, G. (1997). Concurrent verbalisation during drivers' visual search and hazard perception. *In Rothengatter, T., and Carbonell, E. (eds), Traffic and Transport Psychology. Theory and application,* 121–135.

De Craen, S., Vissers, J. A. M. M., Houtenbos, M., Twisk, D. A. M. (2005). Young drivers experience: the results of a second phase training on higher order skills: evaluation study in the framework of the European project NovEV. SWOV report 39.401. Leidschendam.

Deffenbacher, K.A. (1983). The influence of arousal on reliability of testimony. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.) *Evaluating Witness Evidence*. John Wiley, Chichester.

Degia, A. (2007). Qualities of the competent trainer. TRL report PPR 284. Wokingham, Berkshire: Transport Research Laboratory (TRL).

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J. M., Nikolic, D., and Manning, C. A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly, 6,* 1–20.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organisation of behaviour. *Psychological Review*, 66, 183–201.

Elander, J., West, R. and French, D. (1993). Behavioural correlates of individual differences in road-traffic crash risk: An examination of methods and findings. Psychological Bulletin, 113, 279–294.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. Proceedings of the Human Factors Society 32nd Annual Meeting (97–101). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors,* 37(1), 65–84.

Endsley, M. R. (2000). Direct measurement of situation awareness: validity and use of SAGAT. In M. R. Endlsey and D. J. Garland (eds.). *Situation Awareness Analysis and Measurement*, 147–175. Lawrence Erlbaum Associates. London.

Endsley, M.R., Selcon, S.J., Hardiman, T.D., Croft, D.G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. Paper presented at the 42nd Annual Meeting of the Human Factors and Ergonomics Society. Chicago, IL. Human Factors Society.

Engstroem, I., Gregersen, N. P., Hernetkoski, K., Keskinen, E., and Nyberg, A. (2003). Young novice drivers, driver education and training: Literature review. VTI report n491A. Linkoeping: Statens Vaeg- Och Transportforskningsinstitut.

Evans, L. (1991). *Traffic safety and the driver.* New York. Van Nostrand Reinhold.

Evans, T. and MacDonald, W. (2002). Novice driver situation awareness and hazard perception: an exploratory study. *Road Safety Research, Policing and Education Conference*, 2002, Adelaide, South Australia, Australia, 1, 85–90.

Farrand, P., and McKenna, F. P. (2001). Risk perception in novice drivers: the relationship between questionnaire measures and response latency. *Transportation Research Part F: Traffic Psychology & Behaviour,* 4(3), 201–212.

Finn, P. and Bragg, B (1986). Perception of risk of an accident by young and older drivers, *Accident Analysis and Prevention* 18, 289–298.

Fisher, D. L., Pollatsek, A. P., and Pradhan, A. (2006). Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injury Prevention,* 12, 25–29.

Fitts, P. M. and Posner, M. I. (1967). *Human Performance*. Belmont, C.A.: Brookes/Cole Publishing Company.

Fitzgerald, E. S., and Harrison, W. A. (1999). Hazard perception and learner drivers: a theoretical discussion and an in-depth survey of driving instructors, Monash University, Accident Research Centre, Wellington, Australia.

Grayson, G. B., and Groeger, J. A., (2000). Risk, hazard perception, and perceived control. Proceedings of the Novice Drivers Conference Held Bristol, UK 1–2 June 2000.

Gregersen, N. P. (1994). Systematic cooperation between driving schools and parents in driver education, an experiment. *Accident Analysis and Prevention*, 26(4), 453–461.

Groeger, J.A. and Brown, I.D. (1989). Assessing ones own and others driving ability–– influences of sex, age, and experience, *Accident Analysis and Prevention* 21(2), 155–168.

Gugerty, L.J. (1997). Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory, *Journal of Experimental Psychology: Applied*, 3, 42–66.

Hatakka, M., Keskinen, E., Gregersen, N.P., Glad, A., Hernetkoski, K. (2002) From control of the vehicle to personal self-control; broadening the perspectives to driver education. Transportation Research Part F. Traffic psychology. 5, 201–215.

Horrey, W. J., Lesch, M. F., and Garabet, A. (2008). Assessing the awareness of performance decrements in distracted drivers. *Accident Analysis and Prevention*, 40, 675–682.

Horrey, W. J., and Wickens, C. D. (2003). Multiple resource modelling of task interference in vehicle control, hazard awareness and in-vehicle task performance. *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.*

Horswill, M. S., and McKenna, F. P. (2004). Drivers' hazard perception ability: situation awareness on the road. In S. Banbury and S. Tremblay (eds.) A cognitive approach to situation awareness: Theory and application. Aldershot; Ashgate Publishing, 155–175.

Horswill, M., Waylen, A. E., and Tofield, M. I. (2004). Drivers' ratings of different components of their own driving skill: A greater illusion of superiority for skills that relate to accident involvement. *Journal of Applied Social Psychology*, 34(1) 177–195.

Hughes, P. K., and Cole, B. L. (1986). What attracts attention when driving?, *Ergonomics*, 29(3), 377–391

Hull, M. (1991). Mandatory hazard perception testing as a means of reducing casualty crashes amongst novice drivers. VTI Report 372A:PT2: 139–164.

Hull, M. A., and Christie, R. J. (1993). The hazard perception test: the Geelong trial and future developments. VicRoads Report GR 93–13.

Jones, M. (2000). Subjective measures of situation awareness. In M. R. Endlsey and D. J. Garland (eds.). *Situation Awareness Analysis and Measurement*, 113–128. Lawrence Erlbaum Associates. London.

Kass, S.J., Cole, K.S., and Stanny, C.J. (2007). Effects of distraction and experience on situation awareness and simulated driving. Transportation Research Part F, 10, 321–329.

Katila, A., Keskinen, E., Hatakka, M. and Laapotti, S. (1996), 'Skid Training for Novice Drivers - Benefit for Adults, Pitfall for Youngsters', Paper presented at the International Conference of Traffic and Transport Psychology, Valencia, Spain.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and Ramsey, D. J. (2006). The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic *Driving Study Data.* Washington DC: National Highway Traffic Safety Administration.

Kuiken, M. J., and Twisk, D. A. M. (2001). Safe driving and the training of calibration. SWOV report 32.122. Leidschendam.

Lajunen, T., Corry, A., Summala, H., and Hartley, L. (1997). Impression management and self-deception in traffic behaviour inventories. Personality and individual differences, 22(3), 341–353.

Lee, Y., Lee, J. D., and Boyle, L. N. (2007). The effect of voice interactions on drivers' guidance of attention. *Proceedings of the Fourth International Driving Symposium on human Factors in Driver Assessment, Training and Vehicle Design.*

Marsh, H. S. (2000). *Beyond Situational Awareness: the Battlespace of the Future.* Washington, D.C.: Office of Naval Research, March, 2000.

Mathews, M. and A. Moran, A. (1986). Age differences in male drivers' perception of accident risk: The role of perceived driving ability, *Accident Analysis and Prevention* 18, 299–313.

Maycock, G. (2002). Novice driver accidents and the driving test. TRL Report TRL527. Wokingham, Berkshire: Transport Research Laboratory (TRL).

Maycock, G., and Forsythe, E. (1997). Cohort study of learner and novice drivers. Part 4: Novice driver accidents in relation to methods of learning to drive, performance in the driving test and self assessed driving ability and behaviour. TRL Report (TRL275). Wokingham, Berkshire: Transport Research Laboratory (TRL).

Maycock, G., Lockwood, C. R., and Lester, J. (1991). The accident liability of car drivers. Research Report RR315. Wokingham, Berkshire: Transport Research Laboratory (TRL).

McCormack, S. (2003). Recent developments in driver training. 68th Road Safety Congress: Safer Driving, Reducing Risks, Crashes and Casualties.

McDaniel, M. A., Hartman, N. S., and Grubb III, W. L. (2003). Situational Judgment Tests, Knowledge, Behavioral Tendency, and Validity: A Meta-Analysis. *Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology*. Orlando.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., and Grubb III, W. L. (2007). Situational judgement tests, response instructions, and validity: a meta-analysis. *Personnel Psychology*, 60, 63–91.

McDaniel, M. A., and Whetzel, D. L., (2005a). Situational judgement test research: informing the debate on practical intelligence theory. *Intelligence*, 33, 515–525.

McDaniel, M. A., and Whetzel, D. L. (2005b). Situational judgement tests. *Presentation at the 29th annual IPMAAC conference on personnel assessment.* June 19–22, 2005, Orlando, Florida.

McDaniel M. A., Whetzel D. L., Schmidt F. L., and Maurer S. D. (1994). The validity of employment interviews: a comprehensive review and meta-analysis. *Journal of applied psychology.* 79 (4), 599–616.

McDaniel, M.A., Whetzel, D. L., Hartman, N. S., Nguyen, N., and Grubb, W. L. (2006). Situational judgement tests: validity and an integrative model. In R Plyhart and J Weekley (eds). *Situational judgement tests: theory, measurement and application*. Jossey Bass. 183–204.

McGowan, A., and Banbury, S. (2004). Evaluating interruption-based techniques using embedded measures of driver anticipation. In: S. Banbury and S. Tremblay (Eds). *A Cognitive Approach to Situation Awareness*, Ashgate, Aldershot, UK.

McKenna, F. P., and Crick, J. L. (1991). Hazard perception in drivers: a methodology for testing and training. TRL Contractor Report, CR313. Wokingham, Berkshire: Transport Research Laboratory (TRL).

McKenna, F. P., and Crick, J. L. (1993). A cognitive-psychological approach to driver training: the use of video technology in developing the hazard perception skills of novice drivers. *Behavioural research in road safety III. Proceedings of a seminar at the university of Kent*, 22–23 September 1993, 56–61

McKenna, F. P., and Farrand, P. (1999). The Role of Automaticity in Driving. In G.B. Grayson (ed.), *Behavioural Research in Road Safety IX*, Wokingham, Berkshire: Transport Research Laboratory (TRL).

McKenna, F. P., and Horswill, M. S. (1999). Hazard perception and its relevance for driver licensing. *IATSS Research*, 23, 36–41.

Merickle, P. M. (1992). Perception without awareness: critical issues. *American Psychologist*, 47, 792–795.

Nederhoff, A. J. (1984). Methods of coping with social desirability bias: a review. European Journal of Social Psychology, 15, 263–280.

Nguyen, N. T., and McDaniel, M. A. (2003). Response instructions and racial differences in a situational judgement test. *Applied H.R.M. Research*, 8(1) 33–44.

Nisbett, R., E., and Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., and Lawrence, A. (2007). Incremental validity of situational judgement tests for task and contextual job performance. *International Journal of Selection and Assessment*, 15(1), 19–29.

Paulhus, D. L. (1991). Measurement and control of response bias. In Robinson, J. P., Shaver, P. R., and Wrightsman, L. E. (eds.), *Measures of personality and social psychological attitudes* (17–59). Sand Diego, CA: Academic Press.

Paulus, D. L., and Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307–317.

Quimby, A. R., Maycock, G., Carter, L. D., Dixon, R., and Wall, J. G. (1986). Perceptual abilities of accident involved drivers. Research Report RR27. Wokingham, Berkshire: Transport Research Laboratory (TRL).

Rahman, M. (2007). Tackling situational judgement tests. *British Medial Journal Careers.* May 2007, 189–190.

Redelmeier, D. A., and Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine*, 336(7), 453–458.

Roads and Traffic Authority, NSW, Australia (2001). *Hazard Perception Handbook.*

Roediger, H. L., and McDermott, K. B. (1993). Implicit memory in normal human subjects. In Boller F, and Grafman, J. (eds), *Handbook of Neuropsychology*, 8, 63–161.

Rowe, R. M. (1997). *Anticipation in Skilled Performance*. Unpublished PhD thesis. University of Reading, Reading, UK.

Sackeim, H. A., and Gur, R. C. (1979). Self-deception, other-deception, and self-reported psychopathology. Journal of Consulting and Clinical Psychology, 47, 213–215.

Sexton, B. (2000). Hazard perception testing. *Proceedings of the Novice Drivers Conference*, Bristol, UK, 1–2 June 2000.

Shanks, D. R. & St. John, M. F (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17, 367–447.

Shinar, D. (1978). *Psychology on the road— the human Factor in traffic safety*. New York: Wiley.

Siegrist, S. (ed) (1999). Driver training, testing and licensing—towards theory-based management of young drivers' injury risk in road traffic. Results of EU-project GADGET. Guarding Automobile Drivers through Guidance, Education and Technology. Work Package 3. Bfu report 40, Berne.

Smith, B. L., and Murdoch, D. K. (1986). Commentary driving applied to safety evaluation of low-volume rural roads: training and use. *Transportation Research Record* (1055): 34–44.

Soliday, S. (1975). Development and preliminary testing of a driver hazard questionnaire. *Perceptual and Motor Skills*, 41(3), 763–770.

Spence, C., and Read, L. (2003). Speech shadowing while driving: On the difficulty of splitting attention between eye and ear. *Psychological Science*, 14(3), 251–256.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., et al. (2000). *Practical Intelligence in Everyday Life.* New York. Cambridge University Press.

Stanton, N. A., Chambers, P. R. G., and Piggott, J. (2001). Situational awareness and safety. *Safety Science*, 39, 189–204.

Strauss, R., and Kirlik, A. (2006). Situation awareness as judgement II: experimental demonstration. *International Journal of Industrial Ergonomics*, 36, 475–284.

SWOV (2007). *SWOV fact sheet.* SWOV, Leidschendam, Netherlands, September 2007.

Taylor, R. M. (1990). Situation Awareness Rating Technique (SART): the development of a tool for aircrew systems design. Paper 3 in: *Situational Awareness in Aerospace Operations, AGARD-CP-478.* Neuilly-sur-Seine, France: NATO-AGARD. 3/1–3/17.

Taylor, R. M., and Selcon, S. J. (1991). Subjective measurement of situational awareness. In: *`Designing for Everyone', Proceedings of the 11th Congress of the International Ergonomics Association (Paris),* 789–794.

Taylor, S. E., and Bornw, J. D. (1988). Illusion and well-being: a social psychologist perspective on mental health. Psychological Bulletin, 103, 193–210.

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver, *Ergonomics*, 50(8), 1235–1249.

Underwood, G., Crundall, D., and Chapman, P. (2002). Selective searching while driving: the role of experience in hazard detection and general surveillance. *Ergonomics*, 45(1), 1–12.

Verkasalo, M., and Lindeman, M. (1994). Personal ideals and socially desirable responding. European Journal of Personality, 8, 358–393.

Vissers, J., Mesken, J., Roelofs, E., and Claesen, R. (2008). New elements in the Dutch practical driving test: a pilot study. In Dorn, L (ed.), *Driver behaviour and training*, Vol III, 37–50.

Wallis T. S. A, and Horswill M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident Analysis and Prevention*, 39, 1177–1185.

Wells, P., Tong, S. J., Sexton, B., Grayson, G. and Jones, E. (in press). *Cohort II: a study of new drivers.* Wokingham, Berkshire: Transport Research Laboratory (TRL).

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and D.R. Davies (Eds.), *Varieties of attention.* (pp. 63–102). New York, NY: Academic Press.

Wilde, G. J. S, (1993). Improve risk perception and reduce risk acceptance: two proposals for driver education. *Proceedings of the novice driver education symposium.* University of Alberta, April 22–23, 29–43.

# Appendix A The Gadget or 'GDE' Matrix—reproduced from CIECA (2007)

## GDE matrix: Goals for Driver Education

*Hatakka, Keskinen, Glad, Gregersen, Hernetkoski, 2002*

| | | Essential elements of driver training | | |
|---|---|---|---|---|
| | | Knowledge and skills | Risk-increasing factors | Self-evaluation |
| Hierarchical levels of driver behaviour | Personal characteristics, ambitions and competencies | • lifestyle<br>• peer group norms<br>• personal values and norms<br>• etc. | • sensation-seeking<br>• adapting to social pressure | • impulse control<br>• risky tendencies<br>• personal risky characteristics |
| | Trip-related context and considerations | • choice of route<br>• estimated driving time<br>• estimating urgency of the trip | • physiological condition of driver<br>• social context and company in vehicle | • personal skills with regard to planning<br>• typically risky motives when driving |
| | Mastery of traffic situations | • application of traffic rules<br>• observation and use of signals<br>• anticipation of events | • vulneable road users<br>• breaking traffic rules / unpredictable behaviour<br>• information overload<br>• difficult (road) conditions | • strengths and weaknesses regarding driving skills in traffic<br>• personal driving style |
| | Basic vehicle control | • control of direction and position of car<br>• technical aspects of vehicle | • improper use of seatbelt, headrest, sitting position<br>• under-pressure tyres | • strengths and weaknesses of basic vehicle control |

# Appendix B Methods used—description

In this section, we outline the methods used in the review, and give details of their implementation.

## B.1 Methods

The following methods were used to gather relevant information for the review:

- formal searches of literature databases (see below for search terms used);
- informal internet searches through Google and Google Scholar;
- reference to existing relevant major European projects;
- semi-structured interviews with GB police driving instructors.

## B.2 Databases searched

The following databases were used to search for relevant publications:

- **TRL Knowledge Base:** The TRL Knowledge Base comprises a number of databases including the Transport Research Abstracting & Cataloguing System (TRACS). This is the main catalogue of publications held both in the TRL library and elsewhere. It contains bibliographic references and abstracts of English and foreign language articles from journals, books and research reports. It is the English language version of the worldwide ITRD database (International Transport Research Documentation) and contains abstracts from publications in the USA, Australia, Scandinavia, the Netherlands and Canada in addition to UK material. The database has been updated daily since 1972 and now comprises 260,000 items. This is the prime literature resource for transport research. The Knowledge Base also includes the PROJEX database that contains summaries of current and recently completed research projects undertaken in ITRD member countries.
- **Science Direct:** Science Direct provides access to 1500 Elsevier journals and the full text of 14 key transport titles.
- **IngentaConnect:** IngentaConnect offers access to 30,000 publications, including 10,000 online.

## B.3 Search terms

The following search terms were used to search the databases listed in Section B.2:

- Situational judgement; situational judgement assessment/ measurement/ tests/ measures/ scales/ inventories/ screening/ training; situation judgement; situation judgement assessment/ measurement/ tests/ measures/ scales/ inventories/ screening/ training;
- Situational awareness; situation awareness; SA;
- Hazard awareness/ perception/ prioritisation/ management;
- Concurrent verbalisation; delayed verbalisation; commentary driving; commentary driving training/ debrief/ training manuals; think aloud technique;
- Driver workload, dual task performance; mental resources and performance; performance under stress; task intrusion; task switching;
- Skill acquisition; cognitive control;

For the ITRD database, terms were selected by looking at the free text terms above and mapping them against the most similar term in the thesaurus held by TRL for that database, and these 'thesaurus' terms and the free text terms were both used in the final search. For other databases, the free text terms above were used as listed. These free text search terms were also used informally in the internet search engines Google, and Google Scholar.

Items were selected for inclusion in the resulting list of 'possible' references if any of the terms appeared in the title, abstract, or keywords of the search databases. References returned through the searches were then analysed, and an audit trail was kept of decisions to include or exclude each reference from the review. A reference was included directly in the review (i.e. it was cited) if at least one of the following was true:

1. It provided direct evidence addressing the use of situational judgement and situational awareness measures in driving research or training and assessment.
2. It was necessary either alone or as part of an extant literature for illustration of the key concepts of: situational judgment; situation awareness; hazard perception; commentary driving; or other behavioural factors related to accident risk, driving performance, or driver behaviour.
3. It was deemed relevant in some other way by the first author.

Other references, although not included, were perused for relevant cited literature in their reference sections that met one of the three criteria listed above. The final reports from the major European studies listed in Section B.4 were also subjected to the criteria for inclusion.

## B.4 Other relevant projects

Major European studies relating to driver training and testing were identified as being potentially relevant to the issues of situational judgement and commentary driving in driver training and assessment, and were included in the review.

### *B.4.1 TEST*

The TEST project mainly focused on the driving test; it aimed to define the relationship between how long a test lasts, where the test takes place and what is dealt with in the test. The project also aimed to examine whether the driving tests conducted in different countries in Europe or in different test centres in the same country are comparable, and if they require the same skills and attitudes from the candidates. The final objective was to discover whether current driving tests are suitable for dealing with the problem of high accident rates among novice drivers. To do this, it discussed the training needs of drivers as structured by the GDE framework, and suggested future developments in testing that might be used to encourage improved training, covering the whole of the GDE framework.

### *B.4.2 BASIC*

The objective of the BASIC project was to enhance traffic safety by providing comprehensive information concerning new basic training methods and their possibilities to decrease accident risk amongst drivers. Therefore, a description of new basic driver training methods applied in EU and an analysis concerning factors of the various models which are most effective were elaborated.

### *B.4.3 ADVANCED*

ADVANCED was a study of post-licence driver and rider training co-financed by the European Commission. It described and analysed voluntary post-licence training. It also made a series of recommendations on how to improve such training. The report described European research into post-licence training and current knowledge on adult learning. The project also described the typical types of courses available across Europe in terms of programme content, methods used, trainer characteristics and other forms of quality assurance. The recommendations section offers advice on how to make fundamental improvements to post-test training and its delivery.

### *B.4.4 HERMES*

The three year EU HERMES project began in March 2007 and focuses on the development of an easy-to-use training package for driving teachers on the subject of 'how to coach'. HERMES draws on existing experience of coaching methods in driver training and on expert advice in the greater coaching world. The envisaged end result is a training package which new and experienced driving teachers can follow, including:

1. What is coaching—the principles of coaching.
2. Why coach—aims and rationale of coaching in driver training.
3. When—and when not—to coach: practical application in driver training.
4. How to coach—methods to be used by the driving teachers.
5. Scenarios for coaching: training exercises.

Other relevant CIECA publications, such as the report on integrating the GDE matrix into category B driver training and the practical driving test were also included.

## B.5 Semi-structured interviews with GB police driving instructors

A number of police driving instructors were contacted and interviewed regarding their views and experience specifically of commentary driving, but also of situational-based measures in advanced driver training. Their views on the use of such techniques for learner drivers were included in the review.

# Appendix C Additional example SJT items

## C.1 SJT item taken from Rahman (2007)

*You have just started a job as a medical F2 in a new hospital. Your wife has a chest infection, and is not yet registered with a GP and has asked you to prescribe antibiotics. Rank the following options 1–5, 1 being the most effective/best option, five being the least effective/worst option:*

*Prescribe the medication as a private prescription, and arrange for her to register with a GP the following week.*

*Tell her to register with a GP locally.*

*Prescribe the medication on a hospital take home prescription with her details on it.*

*Prescribe the medication on a hospital take home prescription with one of your patient's details on it. Collect the medication from the hospital pharmacy.*

*Ask one of your work colleagues to write a prescription on a hospital take home script without seeing your wife.*

Question taken from the Emedica online revision for GP ST Stage 2 assessment. Answers and explanations at www.emedica.co.uk/bmjsjt.htm

## C.2 SJT item taken from McDaniel and Whetzel (2005a)

*You assigned a very high profile project to one of your project managers. The project is very complex and involves the coordination of several other project managers. During each of the project update meetings, your project manager indicates that everything is going as scheduled. Now, one week before the project is due, your project manager informs you that the project is less than 50% complete.*

***Candidates were asked to rate the effectiveness of the following behaviour responses in resolving the problem—note McDaniel and Whetzel (2005) did not print the exact instructions>***

*Personally take over the project and meet with the customer to determine critical requirements.*

*Meet with the customer to extend the deadline. Talk with the project manager about how the lack of communication has jeopardized the company's relationship with the customer.*

*Fire the project manager and take over the project yourself.*

*Coach the project manager on how to handle the project more efficiently.*

*Do not assign any high profile jobs to this project manager in the future.*

## C.3 SJT item taken from McDaniel and Whetzel (2005b)

Everyone in your work group has received a new computer except you. What would you do?

1. *Assume it was a mistake and speak to your supervisor.*
2. *Confront your supervisor regarding why you are being treated unfairly.*
3. *Take a new computer from a co-worker's desk.*
4. *Complain to human resources.*
5. *Quit.*

# Appendix D Questions asked in Police driving instructor interviews

The following questions were asked in interviews with two police driving instructors.

**Question 1: How is commentary driving integrated into the driver training process? That is, when does it happen in the curriculum?**

**Question 2: Can you describe the process of teaching commentary driving to police trainees?**

**Question 3: How does the trainee know what he is expected to do in commentary driving?**

**Question 4: What materials are used in the training?**

**Question 5: What does the trainer explain about commentary driving?**

**Question 6: Is there a formal test of commentary driving ability in the police training?**

**Question 7: If so, what do trainees need to achieve to pass it?**

**Question 8: What are the criteria for failing the assessment?**

**Question 8a (prompt): What kinds of things would prompt you to say the trainee 'needed more work'?**

**Question 9: What length of training time is normally required for an officer to reach the required standard in commentary driving?**

**Question 10: What in your experience are pre-requisites for commentary driving (e.g. sufficient proficiency in independent driving etc.)?**

**Question 11: What are the skills required to be a good commentary driving trainer?**

**Question 12: What training do trainers for commentary driving need/receive in the police force?**

**Question 13: In your experience, are there differences between good and bad trainers for commentary driving?**

**Question 14: If so, what do you think are the factors that make them particularly good/ particularly bad?**

**Question 15: How do you assess the quality of commentary driving?**

**Question 16: How do you differentiate a good response from a bad response?**

**Question 17: What are the characteristics of a good response, what are the characteristics of a bad response? Please explain using examples.**

**Question 18: How variable are drivers in the commentary driving performance? Are there some who never learn to do it?**

**Question 19: Are there differences in the speed with which trainees pick up commentary driving skills? If so, what do you think are the reasons for this?**

**Question 20: What are your recommendations for applying commentary driving to the training of learner drivers?**

**Question 21: What do you think will be the minimum pre-requisites for the use of commentary driving by DSA in the training and testing of learner drivers?**

# Appendix E Questionnaire given to Police driving trainees

The following questionnaire was distributed to several Police driver trainees at a Police driver training school in the South of England.

**Questionnaire on commentary driving:**

**Background:** TRL (Transport Research Laboratory) have been asked to find out about 'commentary driving', the technique used in Standard and Advanced Police driver training. The reason is to see whether it would be suitable for use with drivers other than the Police (e.g. learner drivers). If you would like to take part in this study, please answer the following questions as openly and as honestly as possible. Any information you give is completely anonymous, and we will not report any of your opinions in a way that they can be attributed to you as an individual.

If you have not yet done any commentary driving, the please fill in the personal details, and then answer only question1. If you have done commentary driving, then please answer all questions.

**Personal Details:**

| | |
|---|---|
| **Age:** | |
| **Gender:** | |
| **Years driving experience:** | |
| **Stage in Police driver training (e.g. Basic, Standard, Advanced):** | |
| **Amount of commentary driving experience so far (approx hours using the technique while driving yourself):** | |

**Questions:**

| | |
|---|---|
| **1: If you have not yet done any commentary driving as part of your training—what are your anticipated general thoughts on the technique? (If you have done commentary driving, then try to remember What you thought before you did any).** | |

| | |
|---|---|
| | |
| **2: If you have done commentary driving, what are your general thoughts on the technique now you have done it?** | |

| | |
|---|---|
| **3: On a scale of 1 to 5 (where 1 is very easy, and 5 is very difficult) how challenging did you find the technique when you first tried it?** | |

| | |
|---|---|
| **4: On a scale of 1 to 5 (where 1 is very easy, and 5 is very difficult) how challenging do you find the technique now?** | |

| | |
|---|---|
| **5: When you first tried the technique, how do you feel it impacted on your general driving, if at all (positively or negatively)?** | |

| | |
|---|---|
| **6: When you use the technique now, how do you feel it impacts on your general driving, if at all (positively or negatively)?** | |

| | |
|---|---|
| **7: When you first used the technique, what did you find most challenging about it?** | |

| | |
|---|---|
| **8: When you use the technique now, what do you find most challenging about it?** | |

| | |
|---|---|
| **9: Do you have any other points you would like to make about commentary driving?** | |

**Thank you for your time. Please return this questionnaire to:**

Dr. Shaun Helman
Email: shelman@trl.co.uk

TRL
Crowthorne House
Nine Mile Ride
Wokingham
Berkshire
RG40 3GA
United Kingdom

Direct: +44 (0)1344 770650
Mobile: +44 (0) 7968 026445

# Situational judgement in driver training and assessment: a literature review

Situational judgement testing and commentary driving are two methods that can be used to assess the higher cognitive processes that underlie observed driving behaviours. The research literature is reviewed, and advice sought from expert police drivers, to assess the suitability of these two methods for use in practical driver training and testing in Great Britain (GB). First the terms "situational judgement" and "commentary driving" are defined within the driving context. Situational judgement is defined as "the application of driving-related knowledge and behavioural tendencies (e.g. personality, attitudes, beliefs etc.) to the assessment of traffic situations, including knowledge of the appropriate driving behaviours in those situations", and is distinguished from the related terms of "hazard perception" and "situation awareness". Commentary driving is defined as "the verbalisation of at least some of the driving-related contents of awareness, while actually driving through a situation". Key findings from the relevant literatures – related to the practical use of these methods in the GB driver training and testing protocol – are then listed. It is concluded that neither method is suitable as a "pass-fail" instrument in the practical driving test in GB, although situational judgement testing has considerable promise as a method for use in the driving theory test, as long as good-practice procedures for test creation are followed. The focus of situational judgement testing should be on the assessment of safe "behavioural tendencies" in driving as well as "knowledge" of what constitutes safe and appropriate behaviour. It is suggested that commentary driving would not be suitable for learner drivers while they are in control of the vehicle, due to possible deleterious effects on driving performance.

**Other titles from this subject area**

**PPR096** The Heavy Vehicle Crash Injury Study (HVCIS) Project Report. I Knight, R Minton, P Massie, T Smith and R Gard. 2008

**PPR248** Review of International Road Safety Good Practice. J A Castle and G E Kamya-Lukoda. 2007

**PPR247** Review of Road Safety Good Practice in English Local Authorities. J A Castle and G E Kamya-Lukoda. 2007

**PPR242** Reporting of road traffic accidents in London: Matching Police STATS19 with hospital accident and emergency data. Supplementary report for St. Thomas' Hospital Central London. H Ward, S Robertson, K Townley and A Pedler. 2007

**PPR223** New and improved accident reconstruction techniques for modern vehicles equipped with ESC systems. R F Lambourn, P W Jennings, I Knight and T Brightman. 2007

**PPR214** SCOTSIM: An evaluation of the effectiveness of two truck simulators for professional driver training. N Reed, A M Parkes, C Peacock, B Lang and L Rehm. 2007

Price code: 3X

ISSN 0968-4093

ISBN 978-1-84608-821-6

PPR312

**TRL**
Crowthorne House, Nine Mile Ride
Wokingham, Berkshire RG40 3GA
United Kingdom
T: +44 (0) 1344 773131
F: +44 (0) 1344 770356
E: enquiries@trl.co.uk
W: www.trl.co.uk

Published by

**IHS**
Willoughby Road, Bracknell
Berkshire RG12 8FB
United Kingdom
T: +44 (0) 1344 328038
F: +44 (0) 1344 328005
E: trl@ihs.com
W: http://emeastore.ihs.com