Development of hazard perception testing

Hazard Perception Testing

Barry Sexton
Transport Research Laboratory

ABSTRACT

This paper discusses the background to hazard perception testing and in particular the relationship with accident liability. It then summarises the research tasks undertaken in a recent project to evaluate hazard perception testing of learner drivers, with the objective of it being introduced to the driver licensing system.

A small trial of items selected from hazard perception tests produced by NFER (National Foundation for Educational Research), during the work on the L-driver Theory Test was used to evaluate the performance of novice and experienced drivers. Subjects’ accident liabilities were modelled for age, mileage and experience as well as the hazard perception score. The results were encouraging in that a relationship with accident liability was indicated, albeit not statistically significant.

New items were filmed using staged situations as well as opportunistic filming. Filming produced sufficient items to be evaluated in a trial on groups of people with different levels of driving experience. The results of this trial found that 38 items trialled had the required psychometric characteristics to create two new hazard perception tests.

While this test development was being undertaken, a specification for suitable hazard perception training material was developed. The resultant specification was given to DSA who then created a training package with three video-based interactive training modules.

A central part of this work was a trial of the trainability of hazard perception skills in learner drivers. Candidates were tested on two occasions and training was given before the second test. The gain in hazard perception scores relative to an untrained group was analysed to determine the benefits of training.

1. BACKGROUND

Research at the Transport Research Laboratory has shown that hazard perception skills relate to potential for crashes, especially for inexperienced drivers, (Quimby et al 1981, Maycock et al 1991). There is also some evidence that training can improve hazard perception skills (McKenna et al 1994) and so may help to reduce accident liability.

The following extracts from a range of published papers illustrate that hazard perception is related to accident liability, and that drivers with better hazard perception skills have lower accident rates. Further, it has also been shown that hazard perception skills can be reliably taught to novice drivers.
**Early TRL findings** - a positive correlation of hazard perception with self-reported accidents, even after adjusting for exposure:

"... Having fitted a basic model which included miles per year and age, the other measured driver abilities were tested for correlation (with self-reported accidents) by adding them singly to this model. ..... Median latency is a measure obtained in the simulator; it is an appropriate measure of the time a driver takes to respond to hazards .... The correlation shows that the longer the subject's response time the higher their accident frequency" (Quimby et al 1986)

**McKenna statements**

"The decrease in response latencies on retest shown by the group of drivers taking the ROSPA course in advanced driving clearly points to the benefits of such training for hazard perception skills. ... It should be noted that the course run by ROSPA is not solely concerned with the training of hazard perception skills... " (McKenna et al 1994)

"...a question arises as to whether there is any relationship between hazard perception and accident involvement. In our laboratory we have found such a relationship, where drivers who had been involved in an accident in the previous three years had significantly worse hazard perception scores than drivers who had remained accident-free .... " (McKenna et al 1999)

**VicRoads results**

"... Analysis of the Geelong trial data indicated differences in score on the Hazard Perception Test between novices and experienced drivers and between subjects reporting crash involvement and those not crash involved. These differences in score were statistically significant across age groups. The Hazard Perception Test appears to offer discrimination between subjects on the basis of driving experience and recent crash involvement. " (Hull et al 1993)

**ACER report** (Evaluating the hazard perception test introduced into the State of Victoria driver licence system)

"... The measures of hazard perception produced by the HPT have shown that they contribute to identifying those cases involved in some types of police reported casualty accidents for the population in this study. The HPT measures were evident in identifying fatal and serious injury accident for novice drivers generally and identifying other injury accidents for 18 year-old females specifically. ...... If the latent trait that this test is assessing can be measured with more precision, by replacing poor performance items and/or adding more items, the test's predictive powers are likely to increase. ..." (Congdon 1999)
Research has also shown that young drivers have high accident rates during their first year of driving, but that these decrease over the next two years of driving, (Forsyth et al 1995). This is illustrated for new female drivers in Figure 1. (A similar pattern exists for males).

TRL started work on a moving hazard perception test in 1981. This was accomplished by using the front of a real car with a 35mm back projected screen in front of the vehicle. Subjects would use a lever to indicate when they spotted a hazardous scenario. This system evolved into a video version using a computer and an automatic scoring system. By 1994 the MPEG1 video compression technology was available. When MPEG2 became available in 1997, full-screen video became available and with it better clarity of the images.

2. OBJECTIVES OF RECENT TRL RESEARCH

The overall purpose of creating an automated HP prototype test was to evaluate the potential for including a hazard perception test in the theory test. To achieve this purpose several steps were defined. These included:

- producing a design for hazard perception items. These items had to able to distinguish learner, novice and experienced drivers.
- producing a number of sample test items.
- determining if hazard perception scores could be improved with additional training and how much and what kind of training would be required.
- recommending whether hazard perception testing should be part of the UK theory test, the pass level and number of items, etc.

The first part of developing a hazard perception test was to define hazards. Driver trainers often see all things on the road as hazards, a hazard in this context was something that experienced drivers would spot early on and take appropriate preventative driving actions. This means that static (bends, roundabouts) and environmental situations (ice, rain) were excluded, leaving only dynamic situations that require scanning and anticipation or need "keeping an eye on". Dynamic situations are those that involve other road users, such as, other vehicles, pedestrians and cyclists and may well incorporate hazardous static or environmental features.
3. HAZARD PERCEPTION ITEM DEVELOPMENT

Four hazard perception tests were developed by the NFER (National Foundation for Educational Research), under contract to TRL during 1995. These hazard perception tests were broad-based by design, in that they covered a range of driving situations and potentially hazardous scenarios. The importance of these four tests, for the recent work, was that they provided a pool of hazard perception items and that many of the items discriminated between novice and experienced drivers.

An early task in the recent research was to identify 'good' hazard perception clips and see how they performed. Those drivers with 'good' hazard perception would have good scanning skills and good anticipation. 'Good' video clips were thus defined as those requiring drivers to demonstrate these characteristics. It also seemed sensible that a hazard should eventually become obvious to even the least perceptive novice drivers, and that there should be a minimum of 'clutter' leading up to the hazard. Hence the following criteria were defined for suitable hazard events; (these criteria have been dubbed a 'blueprint' for items):

**Characteristics of suitable events**

The event should

- develop into an 'actual hazard'
- permit anticipation for an experienced driver or trained novice
- require scanning ahead and/or to the side
- be a clear and uncluttered scenario
- not just depend upon reaction time

Existing video clips were viewed independently by three TRL 'experts' who were familiar with the earlier TRL work and the more recent NFER work on hazard perception. They applied the above criteria to decide which items should be included in a validation trial.

4. TRIAL OF GOOD HAZARD PERCEPTION DEVELOPMENTS

Novice drivers (<2 years of driving experience) and experienced drivers (>10 years of driving experience) were recruited to take a test constructed from the 13 'best' clips identified from the NFER hazard perception tests. Approximately 50 male and 50 female drivers in each experience group were tested.

Subjects took the test on a PC. They were given instructions on the screen and then shown a practice video clip. The test consisted of a number of clips one after another. Subjects had to respond by pressing a button as soon as they identified a hazard. Scores were given according to when the subject responded within a pre-determined scoring window. If no response was given, the score was zero.

The next three illustrations show an example of how a hazardous scenario develops.
The situation is developing and experienced drivers may have noticed the cyclist.

The cyclist is now clear, as is an approaching car. A hazardous situation has developed, and the test taker should have responded.
It is now too late to respond; the hazardous situation has fully developed.

There were 22 hazard perception events within 13 clips that looked as if they met the 'blueprint' criteria. An event is defined as a hazardous situation within a video clip and is also referred to as an item. A 'good' event or item is one that distinguishes between the different experience groups. Within the 22 events, 16 had a statistically significant difference in mean response times between the experience groups. The mean overall scores based on the 16 items show a statistically significant difference between experienced and novice subjects.

Analysis of self-reported accident data showed that accident models as seen in other studies, (Forsyth et al 1995), for age, experience and mileage were supported by the data from this study. The hazard perception score factor (as measured by the best 16 events) was weakly related to accidents such that higher hazard perception scores were associated with lower accident liability. This association was not statistically significant, probably due to the limited sample size, but lends at least some support to the assumption that increasing driver's hazard perception scores will reduce their accidents.

5. NEW TEST MATERIAL

New hazard items were required to devise new tests for use in a trial of learner driver hazard perception training. The first stage was to film some new clips and then run a trial to see which items could distinguish between drivers of different experience.

Filming of new hazard situations was undertaken by the TRL video media film crew using a left-hand drive car and filming from the driver's eyepoint. DSA Training Establishment staff gave valuable assistance with some of the filming. The film was broadcast quality such that it could be encoded to MPEG2 standard for displaying via a PC. Some panning of the camera was used to good effect on corners and roundabouts. Two main types of filming were undertaken. The first was opportunistic on a variety of road types, and yielded some useful video footage that was included in the subsequent trials. However, it was not very efficient as a process. The second type of filming used staged events and this was generally more efficient.
TRL experts and Steering Group members selected suitable items for trialing. They applied the criteria identified as the 'blueprint' together with their own experience and knowledge of the type of hazard that may be appropriate.

The items were split into two sets, each with about 27 items in each. One item was selected as an example item in order to give subjects an indication of what was expected. The trial required volunteers to take both sets of hazard perception items. Subjects were volunteers and were either learners, novice drivers with less than 2 years experience or experienced drivers who had been driving at least 10 years. The total target sample was 450 subjects with 150 from each group.

The scoring method was empirically based on the distribution of response times. The score given for each hazardous event depended upon the subject's first response within a time-window. The time-window is the time from when the hazard could have first been spotted until it is a fully developed hazard. The earlier within the window that the candidate responds, the higher the score. Candidates who failed to respond in the time-window were given a zero score. The following chart, Figure 2 shows the response distribution for a particular item. In this example there were three main hazardous situations to which subjects were responding, this is reflected in the three response peaks.

**Figure 2. Response frequencies for item 1**

<table>
<thead>
<tr>
<th>Item 1 frequency by time</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

Items were selected to provide two 'new' tests, referred to as X and Y. The following criteria were applied:

- the item should, at least, distinguish between the experienced drivers and other drivers
- the test should take about 15 minutes to complete
- the test should be internally consistent - i.e. the items within the test should be measuring the same thing and someone scoring well on one item is likely to score well on another item
• the two tests should be parallel - i.e. they should be of similar overall difficulty with a similar mix of items
• replicates of the same staged events should not be in the same test

There were 38 'best' items that were able to distinguish between drivers of different experience. Sixteen of the 38 items were used to make test X and sixteen different ones to make test Y. The characteristics of tests X and Y given in Table 1.

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>Learners</th>
<th>Novice (&lt;2yrs)</th>
<th>Experienced (&gt;10yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Mean</td>
<td>24.44</td>
<td>29.32</td>
<td>36.50</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.785</td>
<td>0.819</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td>157</td>
<td>152</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>Reliability = 0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Mean</td>
<td>22.06</td>
<td>27.91</td>
<td>36.54</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.864</td>
<td>0.899</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td>157</td>
<td>152</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>Reliability = 0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mean scores are monotonically increasing with experience and each group mean is statistically significantly different from the others.

An analysis of the hazard perception score showed that it could be predicted by age and average mileage. Older drivers and those with a higher annual mileage had higher scores. The self-reported accident data from a questionnaire showed a significant difference between the average accident liabilities of the novice and experienced drivers.

6. TRAINING MATERIAL

A review of existing driver training materials and discussions with driver training organisations and others suggested that hazard perception could best be taught as part of training in the following three driving skills. Drivers with good hazard perception skills would generally be demonstrating these driving skills:

• keeping safe driving distances
• looking well ahead
• driving at appropriate speed

Training packages were created by DSA Training Establishment at Cardington, to a specification supplied by TRL. The core of the package was filmed traffic scenes. The training had the following characteristics:
candidates received a minimum of one session and a maximum of three sessions of training
each session lasted a maximum of one hour; candidates received one session per week
a session of one hour duration was based upon approximately 20 minutes of video material
the training was delivered in a 'classroom' type setting, in small groups and was interspersed
with their normal on-road tuition

The content of the package facilitated a learning process in which students were required to be
highly active. A method which proved effective was to "freeze" the video, and ask students to
discuss where the hazards were and ask the question 'what if ... happened?’ and to think about what
could happen and what they would do next.

The package was based on video and workbook. Trainers had user guides/manuals, and student
workbooks and other materials. A basic hazard perception training module plus two more advanced
training modules were created.

7. TRIAL OF TRAINING

A trial of the training package was organised using theory test centres across the country.
Candidates who came for their theory test were recruited into one of three groups. Group A, the
control group, were asked to take either test X or test Y hazard perception test and to return about 8
weeks later to take the other test. Group B were asked to take the hazard perception tests as group
A, but were also asked to attend for an hour of basic training between tests. Group C were similar
to group B except that they received one hour of basic training plus two separate hour-long sessions
using the more advanced training modules. It was hoped to obtain about 250 candidates per group, in
practice about 160 per group completed the trial.

Every candidate took one of two hazard perception tests on each test occasion, the order being
counter-balanced. Hence a candidate who took test X on the first occasion took test Y on the
second, and vice versa. Any effect due to the test taken was thus controlled for. The control group
was required because all candidates were exposed to the test prior to any training, and they all
would have been receiving some instruction between taking the two hazard perception tests. It was
therefore necessary to see if their hazard perception changed between the two tests.

Figure 3. Mean difference and 95% confidence interval for gain in HP scores
The results of the training showed that learner driver test score gains increase with the level of training received. The difference in hazard perception score gains, (as measured by the hazard perception tests developed within this project), between groups was statistically significant. The mean values for each group and the 95% confidence interval are shown in Figure 3. The conclusion reached was that hazard perception skills as measured by these tests can be trained in learner drivers.

8. SUMMARY POINTS

The key points from this research programme are:

- A better understanding has been obtained of the type of hazard perception skill that distinguishes between experienced drivers and new drivers
- A definition of the criteria required for hazard perception test items that measure this skill has been derived
- These criteria are referred to as a 'blueprint' and include requirements such as scanning and anticipation
- New hazard perception tests have been developed from events filmed with the 'blueprint' criteria in mind
- These tests are reliable and valid instruments for measuring hazard perception skills and would be suitable for introducing into the driver licensing system
- A group-based hazard perception training package has been defined and developed
- There is clear evidence that hazard perception skills as measured by the newly developed tests can be trained using the training material developed
- The observed gain in hazard perception score from training increases with the level and number of hours trained
- The research has demonstrated the feasibility of introducing hazard perception testing into the driver licensing system
REFERENCES


