EVALUATION OF ROAD SAFETY EDUCATION AND NOVICE DRIVER SAFETY MEASURES IN GREAT BRITAIN

Kate McMahon and Deirdre O'Reilly

Road Safety Division
Department of the Environment, Transport and the Regions (DETR), GB
e-mail: kate_mcmahon@detr.gsi.gov.uk or deirdre_o'reilly@detr.gsi.gov.uk

INTRODUCTION

This paper discusses the issues relating to the evaluation of road safety interventions aimed at improving road user behaviour. By exploring the issues raised through three practical examples we hope to provoke a discussion about the best methods of evaluating the impact and effectiveness of such interventions and in particular to focus on the best outcome measures to be used in assessing behavioural measures. The examples have been selected to encompass a range of types of evaluation (summative and formative; process and outcome) and a range of road users (novice drivers, child cyclists and pedestrians), they are:

- hazard perception testing and training of novice drivers (formative);
- bicycle training for children (summative); and
- practical child pedestrian training for young children (formative).

Developing effective remedial interventions is an iterative process. Formative evaluations are those that are part of the initial development of an intervention. Summative evaluations typically involve an assessment of existing remedial measures implemented on a wider scale either in terms of size or geography.

Ideally evaluation involves assessments of both processes and outcomes. Process evaluation involves assessing the effective delivery of interventions for example what is the mechanism for educating trainers or making the target group aware of the intervention? To some extent this aspect of the assessment reveals the context in which the intervention takes place. Outcome evaluation involves assessing the impact of the intervention on the desired outcomes for example the impact on skills, knowledge, behaviour, accident involvement etc.

While the desired outcome is clearly an improvement in safety, it is not always possible to measure impacts in these terms and we need to consider alternative outcome measures that satisfy the need to objectively test the impact of an intervention on a specified group. In terms of educational interventions aimed improving the safe behaviour of road users it is very hard to clearly demonstrate changes in safety as a result of the intervention. Table 1 gives a simplistic sequential idea of the desired effects:
Table 1:

<table>
<thead>
<tr>
<th>Safety Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowing facts, attitudes and skills</td>
</tr>
<tr>
<td>Understanding</td>
</tr>
<tr>
<td>Behavioural change</td>
</tr>
<tr>
<td>Safety improvement</td>
</tr>
</tbody>
</table>

EVALUATION OF HAZARD PERCEPTION TESTING

Background

Research at the Transport Research Laboratory and at Reading University during the 1980s and early 1990s showed that hazard perception skills appear to be correlated with driver experience and accident liability. This research also showed that performance in a hazard perception can be improved by training. The research used video based material to assess ability to recognise hazards, and found significant differences in performance between inexperienced drivers, experienced drivers and “experts” e.g. police drivers.

A test of hazard perception was therefore recognised as having potential to reduce the relatively high accident rate associated with novice drivers. When the theory test was being developed in GB, the possibility of including hazard perception was considered. In the event, the first generation theory test was not computer based, and therefore any test of hazard perception, which requires moving images, could not be included. The issue of computerisation was reappraised in connection with renewal of the contract for operating the theory test and therefore the issue of hazard perception was reconsidered.

The work at TRL and Reading University involved subjects sitting in front of a video based road scene. They were asked to indicate, using a lever or button, as soon as they saw a hazard, and their score depended on how early each hazard was spotted. A hazard, in this context, is defined as something which a driver should keep an eye on because it could lead to an accident situation.

Results of early research

Research on accident liability and on hazard perception has demonstrated that:

i. inexperience rather than age is the most significant factor determining the higher accident rate of novice drivers;

ii. the ability to recognise hazards as they develop improves with driving experience;

iii. experienced drivers both perform better on hazard perception tests and have fewer accidents;

iv. people can be trained in hazard perception in such a way that they do better at the test.

v. drivers who perform better on HP tests also make better responses to hazards on the road, as assessed by driving instructors.

What past research has not demonstrated is whether drivers whose HP skills, (as measured by these tests), have been improved by training, actually go on to be safer drivers and have fewer accidents.
Development and Evaluation of Hazard Perception Tests and Training

Although earlier research had shown potential for safety benefits, it was felt that further development and evaluation was required before a hazard perception test could be introduced nationally. The tests that had been developed for these research studies were not suitable for implementation. Several matched versions of tests with demonstrated reliability and validity would be required for national testing, and appropriate scoring methods and a pass mark would have to be established.

Consequently, a two-phase project was commissioned from TRL, in 1997. Phase 1 would update knowledge about hazard perception testing; develop a set of tests; devise a training package; and carry out a pilot evaluation. Phase 2 would be designed to provide evidence of the safety benefits of a hazard perception test. It would aim to establish whether people who have been trained to pass the hazard perception test would in practice be safer drivers on the road.

In Phase 2, a sample of learner drivers would be tested on the hazard perception tests developed in Phase 1. Half the sample would be trained in hazard perception and then retested; the remainder would act as a control group. All drivers would be followed up for one year after passing the driving test in order to get information about their subsequent accidents. The objective would be to demonstrate that drivers having the potential to achieve good accident liability could be identified on the basis of their hazard perception scores, and that those whose HP scores were raised by training would also have a better accident record. If, as expected from earlier research, there were shown to be a correlation between hazard perception scores and accident liability, this would provide justification for introducing the test into the national theory test.

In designing Phase 2, consideration had to be given to the likely size of any reduction in accidents following hazard perception training and testing, in order to estimate the required sample size. It was calculated that a minimum contact sample of 8,200 drivers aged 17-24 years would be needed in order to detect a 5% reduction in accident rate, allowing for drop out during the course of the research. This was considered to be the largest sample which could be handled practicably, but it would not be large enough to obtain conclusive results were the accident reduction effect to be less than 5%. This phase of the evaluation project was estimated to cost £800,000 due to the need to train and test such a large sample.

Given the risk of an inconclusive result from Phase 2 and the high cost involved, the need for this full evaluation was reappraised following the firm decision to implement a computer based theory test in 1999. This decision meant that computerisation was not dependent on the results of the hazard perception evaluation. The extra cost required to add a hazard perception test was low – about 75p per test – and even a reduction of 1% in accident risk (too small to be detected in the evaluation project) would provide a return of over 7 times the annual cost of the test (£6.6m compared with £0.9m, per annum).

Early work on Phase 1, reviewing existing hazard perception material, showed that some considerable further development work was required to produce specimen tests for evaluation. This had the consequence that Phase 1 needed to be extended by a year, delaying the completion date of Phase 2, and putting back the date for implementation into the theory test.

It was therefore decided that resources would be better spent on the development and evaluation of hazard perception training material, and on the production of robust tests, which could be evaluated in terms of their ability to discriminate between drivers with different levels of driving experience. An expanded Phase 1 was therefore taken forward, and Phase 2 was cancelled.
This decision meant that the emphasis of the evaluation project was changed. The project was no longer required to produce firm evidence of accident reductions in order to justify introduction of a hazard perception test. The evaluation now focused on

- Improving the quality of the hazard items used in the test;
- Ensuring that the items could be scored consistently in order to distinguish between novice and experienced drivers
- Developing training material;
- Determining if hazard perception scores could be improved with training.

**The revised evaluation project**

The first step was to define hazards. In this context, hazards were considered to be something that experienced drivers would spot early on and take appropriate driving action to prevent. Static hazards such as bends were excluded, and only dynamic situations which require scanning and anticipation or need “keeping an eye on” were included. Dynamic situations involve other road users, and may well incorporate hazardous static or environmental features.

Some material for a hazard perception test was developed as part of the theory test development. A large number of video clips were produced and evaluated for their ability to discriminate between experienced and novice drivers. The performance of these clips varied, and by examining the “best” clips a set of criteria for future filming were drawn up. Clips were needed which required drivers to demonstrate good scanning skills and anticipation. “Good” hazard events 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 on reaction time.

A pilot hazard perception test using the 13 “best” clips from the available material was trialled using 100 drivers with less than 2 years experience, and 100 with more than 10 years experience. The test consisted of a number of clips, one after another. Drivers had to respond by pressing a button as soon as they identified a hazard. Scores were given according to how quickly the driver responded within a pre-determined scoring window. If no response was given, the score was zero.

Within the 13 clips, 22 hazard events appeared to meet the selection criteria. Of these, 16 had a statistically significant difference in mean response times between the two experience groups. Self report accident data were also collected, and the hazard perception score factor, measured on the 16 best events, was weakly related to accidents such that higher hazard perception scores were demonstrated by drivers with lower accident liability.

These results, although not statistically significant due to the small sample size, indicated that the criteria were on the right lines. The next stage was to film new hazard items to produce a set of tests for further evaluation. Staged events were filmed, using a camera inside the car, from the driver’s view point. The material was trialled using learner, novice and experienced drivers with about 150 in each group. The aim of the trial was to produce two equivalent tests, which included the events which best discriminated on the grounds of
experience, which were internally consistent, and which lasted about 15 minutes. Two tests, each containing 16 items were devised, with the following characteristics:

Table 2:

<table>
<thead>
<tr>
<th>TEST</th>
<th>STATISTIC</th>
<th>LEARNERS</th>
<th>NOVICE</th>
<th>EXPERIENCED</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>(Reliability 0.81)</td>
<td>Standard error</td>
<td>0.785</td>
<td>0.819</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>(Reliability 0.84)</td>
<td>Standard error</td>
<td>0.864</td>
<td>0.899</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td>157</td>
<td>152</td>
<td>153</td>
</tr>
</tbody>
</table>

The mean scores are monotonically increasing with experience and each group mean is statistically significantly different from the others. More experienced drivers and those with higher mileage had higher HP scores, and self report accident data showed a significant difference between the accident liabilities of the novice and experienced drivers.

Evaluation of training

The next stage was to devise a training package which could improve drivers' HP scores. A package based on video of driving scenes and a workbook was developed for use in a classroom setting. The package was trialled using learner drivers recruited at the time of their theory test. They were divided into 3 groups: a control group A who were tested on test X or Y, and then eight weeks later on the other test; Group B were tested as A, but also had one hour’s basic training in the period between tests; Group C had two additional one hour more advanced training sessions. All subjects were tested twice, so that they took both tests X and Y.

The results of the training are shown in Table 3, below. The learner driver test scores increased with the level of training received. The difference in HP score gains between groups was statistically significant. The conclusion was that HP skills can be trained in learner drivers.

Table 3

<table>
<thead>
<tr>
<th>TRAINING LEVEL</th>
<th>INCREASE IN TEST SCORE</th>
<th>NONE</th>
<th>BASIC</th>
<th>ADVANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.46</td>
<td>4.63</td>
<td>12.43</td>
<td></td>
</tr>
<tr>
<td>S.E</td>
<td>0.62</td>
<td>0.77</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>160</td>
<td>160</td>
<td>160</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

Full evaluation of the accident reduction benefits of HP testing was judged to be impractical, on the grounds that the large sample required to produce statistically significant results would be difficult and costly to obtain. However, the more limited evaluation produced reliable “blueprint” tests which discriminated significantly on the grounds of experience. The trial of the training package was highly successful.

The decision has now been taken to include HP in the theory test in 2002. A post-implementation evaluation will then be carried out, comparing accident rates of novice drivers before and after the introduction of the new test. Research is also planned to follow a cohort of new drivers through their learner and novice period, to look at the training they have, their test performance, and their subsequent accident rates. This will yield more evidence on the effects of HP testing and training.
EVALUATION OF CYCLE TRAINING

Background

Since the 1950's most local authorities in England and Wales have been providing some form of cycle training for children in their last year of primary school i.e. when they are aged between 10-11 years old (RoSPA, 1993). This is a time when children will be starting to use bicycles as a means of transport rather than a toy and may be looking forward to moving on to secondary school and more independent travel.

There is no single national pedal cycle training scheme and the nature and provision of schemes varies significantly by local authority. It is estimated that approximately a third of each annual cohort of children receive some form of bicycle training before their twelfth birthday. This level of provision represents a significant commitment of resources from Road Safety Officers, trainers and children.

Previous research has suggested that in the short term (immediately after and 3 months later) cycle training is effective in improving skills and knowledge, however, the longer term benefits were, until this study, unquantified. This study was therefore commissioned to assess whether cycling has an effect on road safety knowledge and cycling skills of children aged around twelve (approximately 2 years after they have attended a cycle training course). A secondary objective was to determine whether particular courses are more effective than others.

Evaluation

Cycle training aims to make child cyclists safer on the road by extending their knowledge and cycling skills. Tests of cycling skills focus on two aspects of cycling performance: basic control is concerned with the ability of mount and dismount safely, to ride in a straight line, to go round reasonable curves with falling off and to use the brakes correctly. Safe behaviours include looking for other traffic, signalling and road position, travelling at speeds appropriate to the conditions and observing and obeying road signs. In addition all current schemes include a bicycle safety check and riding practice.

The evaluation was undertaken by means of a case control study. It involved a sample of different types of well established cycle training schemes across a broad geographical area with the intention of selecting a sample of 12-13 year old children in each area. Half the sample had undertaken cycle training and half were not formally trained but all rode a bicycle regularly.

Eight specific schemes were selected to represent the range available including: those that had on-road training and those that had only off-road training; courses administered over different periods of time – a short one week course compared with courses spread out over a number of weeks; and instruction based courses compared with cycling awareness type courses. In each of the eight areas chosen a target of 80 trained and 80 untrained children was set; these were selected from 8 schools in each area. Thirteen trained and untrained children in each school were sampled – this over-sampling allowed for some children not completing the evaluation. The samples were matched by age, sex and cycling experience to enable as consistent a comparison as possible. Schools with a range of educational achievements were selected using Department for Education and Employment data to avoid educational abilities biasing the results.

Each child completed a knowledge quiz, a practical cycling skills test, undertaken by an independent tester, and they were also given a cycling log book to complete. The quiz included questions relating to their knowledge of the Highway Code and also included general questions about their safety behaviour, bicycle and cycle helmet usage, and accident
involvement. The skills test was typically on a quiet local road near the school. For insurance purposes children required parental consent to participate and a risk assessment was undertaken at each site. Basic control, safety and awareness skills were rated. The log book was supposed to provide more data on the amount and type of cycling children undertook over a 7 day period, however, despite a prize draw, compliance with this aspect of the evaluation was very poor and the results were not used.

Results

A total of 1,974 completed the knowledge quiz and 1,566 completed the practical test. Only 807 completed the log book. In the quiz, trained children scored significantly higher than untrained children and overall boys scored significantly higher than girls. In particular trained children answered questions concerning junctions and priority, traffic signs, signals and road marking correctly significantly more than untrained children. For the practical on-road test most (75%) trained children obtained a ‘safe’ rating by the independent assessor as opposed to only half (53%) of untrained children – this is statistically significant. No significant differences were found between boys and girls.

Given the sample sizes and experimental design it was possible to compare the different types of schemes. The four most effective courses with regard to ‘safety rating’ on the practical road-side test were those which included an on-road training element and were conducted over several weeks (rather than intensively over one or two weeks). Also courses that consisted of more than one stage, each stage being completed at a different age were found to be effective. Children who had been trained on cycling awareness type courses (using problem solving approaches) were generally found to perform better than children who had completed the instruction based type courses.

Conclusions

The results strongly suggest that training children to ride safely at around the age of ten has a lasting positive effect on their cycling practice and knowledge of road safety. The tests of knowledge and skills were considered to be proxy’s for actual behaviour. This method of assessment enabled safety knowledge and behaviour to be linked directly to the type of training or not that children had undertaken in a way that observation studies alone could not have achieved.

There are clearly some limitations to this study. It was not feasible to undertake a study of the impact of training on child cycle casualty numbers due to– the expense of collecting hospital data on relatively small samples, the lack of formal records of children who have received cycle training; and the inability to relate hospital casualty data to police accident statistics, due to widespread under-reporting of child cycle accidents. Furthermore the ‘context’ for the impact in terms of children’s exposure as cyclists was not reliably quantified in this study.

EVALUATION OF PRACTICAL ROADSIDE CHILD PEDESTRIAN TRAINING IN DRUMCHAPEL

Background

Research has shown that road crossing is a far from straightforward task requiring the integration of complex perceptual, cognitive and motor skills. Furthermore since children learn from a bottom-up process, i.e. from concrete experiences they develop more structured and strategic approaches, the most effective way to teach them is in realistic situations. This knowledge has led to a shift in child pedestrian road safety education away from rule based verbal instructions to the promotion of practical road-side training. For younger
children pedestrian training should start with practical training – once these skills and understanding have been established there is room to consolidate lessons learned with verbal instructions and classroom based activities.

Despite a good overall road safety record, Great Britain has had a relatively poor child pedestrian safety record when compared to our European counterparts. We know some of the reasons for this and indeed we have been able to identify the characteristics of the ‘highest risk’ groups of children. In particular we know that children from the lowest socio-economic groups (SEG) are five times more likely to be killed when out walking than their highest SEG peers. There are likely to be a numbers of reasons for this – including greater exposure to risk; exposure on faster more risky types of road, less parental supervision etc. It was therefore agreed to participate in a multi-agency study to develop and assess the impact of a community based practical child pedestrian scheme in a deprived housing estate (Drumchapel) located on the outskirts of Glasgow. The scheme was developed by a team of psychologists at the University of Strathclyde and funded by the Regional Authority with support from the local Community Council and the evaluation was funded by DETR.

**Evaluation**

The intervention was a progressive practical road-side scheme involving the training of children aged 5-7 years old in three basic skills: finding a safe place/safe route; crossing between parked cars; and crossing at junctions. The scheme required children to be trained by local volunteers in small groups of two or three using a problem solving, interactive rather than instructional approach. The aim of the scheme was to improve skills and understanding of children and enable them to take an active part in decision making when out and about walking. This was in no way seen as encouraging independent travel for children of this age. The scheme was administered by a local community worker who identified, trained and monitored the local volunteers, sought the support of local schools and managed the scheduling of training sessions.

The evaluation comprised a pre/post case control design in which the skills of children at the road-side were tested and their conceptual understanding of their behaviours were explored through open questions and prompts. A baseline pre-training measure of children’s skills was taken by two independent Road Safety Officers, skills were then tested shortly after training and then a few months later. The skills of a matched group of untrained children were taken at similar intervals.

**Results**

Both trained and untrained children improved in terms of crossing skills as they got older. However, trained children performed significantly better than untrained children on all the specific skills taught, namely: finding safe places and safe routes; developing safe strategies for crossing between parked cars; and crossing safely near intersections. The benefits of training were maintained over a two month period after training ended. The research found that the judgements of trained children appeared to be underpinned by a better conceptual understanding of the task, making them more able to deal with a variety of traffic situations in a flexible way.

It was not possible to assess the effectiveness of adult volunteer trainers individually but taken as a group, the results they achieved were comparable to those achieved by highly qualified staff in earlier studies. This means that unqualified adult volunteers have great potential for use in this method of training.

Not all children were able to participate in the full 6 training sessions required for each skill. It was therefore possible to establish that a minimum of four training sessions was required to improve skills significantly.
There were other less tangible benefits to this training that were reported but not quantified, these included: greater contact between the community and schools; community members involved in solving their own problems; adults getting training and training children can improve confidence and boost morale.

Conclusions

This evaluation suggests that practical child pedestrian training by adult volunteers in small groups at the roadside results in significantly improved skills and conceptual understanding when compared to untrained children of the same age i.e. this type of training can accelerate the natural learning process.

An obvious limitation of this study is that skills and understanding are measured at the roadside under test conditions. To some extent this is adjusted for by comparing trained versus untrained children both before and after the intervention. These tests are therefore assumed to be some sort of proxy for actual behaviour. The research team did attempt to covertly make video observations of children’s actual behaviour at the roadside at the end of the school day but this was met with strong resistance and suspicion by the local community. As with cycle training this method did enable skills and understanding to be related directly to the training level of the child. A further limitation is measuring the longer term impact of the scheme – in terms of benefits to children, trainers and the community and the long term sustainability of the scheme.

DISCUSSION

This section returns to the initial issue of how best to measure the impact of road safety interventions aimed at improving road user behaviour. At one level we want to establish the impact of behavioural interventions on safety – be it accident numbers or rates. However, there are many practical difficulties and costs involved in getting reliable and sufficient samples of casualties and accidents.

Accidents are relatively rare events and we are aware that not all accidents are reported to or recorded by the police. Using alternative data sources (such as hospitals) can be expensive and the question of consistency with police data then arises. In some studies near misses have been used or self reported accident involvement is used – again problems arise with the reliability of such data and in relating them to national statistics.

Furthermore reductions in accidents may not be entirely due to safety improvements or to an individual intervention. Consideration needs, ideally, to be given to exposure – how often, how far or for how long people drive, ride or walk. This is not always easy to measure, especially for vulnerable road users, and much research has been undertaken to improve such tools as travel diaries. In measuring exposure and accident involvement every effort should be made to ensure consistency with routinely collected national data.

As was noted earlier a minimum sample of 8,200 drivers was required to detect a 5% reduction in accident rate. This raised questions as to the practicality of testing such a large sample and following them up over a period of time, as well as the issue of the high costs involved. In the case of the Drumchapel study the area was too small and the intervention too short to detect any statistical variations in the accident rate.

In the past educational interventions have been assessed by measuring knowledge and self reported attitudes. Such approaches did reveal statistical differences between groups and shifts in knowledge and attitudes. But these alone are of limited value and there is much debate in academic circles about how they relate to actual behaviour.

Increasingly evaluations have moved on to include skills testing and the testing of people’s understanding of skills. It can be argued that behaviour cannot be improved unless people
have the necessary skills to do so. Whether they actually deploy them is yet another level of study. Research has focused on understanding skills such as hazard perception, cycling and crossing the road all seemingly automatic functions to the skilled road user. Tools have been developed to measure such skills. These tools are able to discriminate between experts and novices and need to be thoroughly validated. This can be done by extensive testing using experts and novices while controlling for confounding factors such as age, sex, socio-economic group, exposure and experience.

Interventions rarely work in isolation and it is important to record the context within which interventions are tested and also to take into account, as far as practical, confounding factors. The use of control groups can assist in this if measures are applied selectively. Where a measure, such as the hazard perception test, is implemented nationally on a large scale, it should be possible to detect any effect, post implementation, on accident rates by before and after comparisons. This should provide the evidence which ideally would have been the basis of the proposed evaluation.

A good evaluation encompasses a wide range of measures that assess both the context in which the intervention is taking place, the processes of delivery and the outcomes.

REFERENCES


