READING THE RISK OF NEW ZEALAND ROADS:
A COMPARISON OF ACTUAL AND PERCEIVED DRIVING RISK

Prepared for the AA Research Foundation

Samuel G. Charlton, Nicola J. Starkey, John A. Perrone, and Robert B. Isler

Traffic and Road Safety Research Group
School of Psychology
University of Waikato

25 November 2013

This research report, and the work described in it, was performed under contract to The New Zealand Automobile Association Research Foundation, Wellington, New Zealand.

This report should be cited as: Charlton, S. G., Starkey, N. J., Perrone, J. A., and Isler, R. B. (2013). *Reading the Risk of New Zealand Roads: A Comparison of Actual and Perceived Driving Risk.* TARS Research Report. Hamilton: The University of Waikato.

Table of Contents

Executive summary

1. Background 1

2. Research goal and approach 8

3. Driving simulator testing 10

   3.1. Selection of study roads 10

   3.2. Recording test stimuli 10

   3.3. Method 11

      3.3.1. Participants 11

      3.3.2. Apparatus 11

      3.3.3. Procedure 12

   3.4. Results 14

      3.4.1. Comparison of perceived and objective risk 15

      3.4.2. Comparison of video and still photo risk ratings 18

      3.4.3. Consistency of the risk ratings 19

4. Eye-tracking testing 21

   4.1. Method 21

      4.1.1. Participants 21

      4.1.2. Apparatus 21

      4.1.3. Procedure 22

      4.1.4. Analysis 23

   4.2. Results 23

      4.2.1. Comparison of risk ratings 23

      4.2.2. Eye-tracking and fixations for under-rated risks 24

      4.2.3. Analysis of blinks 27

      4.2.4. Analysis of pupil size and saccades 29

5. Naturalistic driving testing 29

   5.1. Method 29

      5.1.1. Participants 29

      5.1.2. Apparatus 30

      5.1.3. Procedure 32

5.2. Results 33
6. Verification testing 34
6.1. Method 34
6.1.1. Participants 34
6.1.2. Apparatus 34
6.1.3. Procedure 35
6.1.4. Analysis 36
6.2. Results 36
7. Discussion 37
8. Acknowledgements 41
9. References 42

List of Figures

Figure 1. An overview of the six research tasks used to answer the research questions. 9

Figure 2. Thumbwheel used by participants to provide risk ratings. 13

Figure 3. A scene from the test video showing the on-screen risk scale. 13

Figure 4. A screen from the still photo recognition/risk rating task. 14

Figure 5. Mean risk ratings for video scenes from four objective risk categories. 16

Figure 6. Mean risk ratings and RPS scores for video scenes. 17

Figure 7. Mean risk ratings and RPS scores for video scenes indicating the types of road situations that were over- and under-rated by the participants. 17

Figure 8. Mean risk ratings for 12 still photos and corresponding video locations. 18

Figure 9. Continuous risk ratings for two viewings of the same section of road. 20

Figure 10. Repeated risk ratings for three viewings of the same photo. 20

Figure 11. The TARS eye-tracking equipment recording a participant's eye movements while viewing a road scene. 22

Figure 12. Mean risk ratings for 20 locations in the eye-tracking video and corresponding locations in the driving simulator test video. 23

Figure 13. Screen positions of a roadside pole used in the analysis of participants' eye-tracking and fixations for under-rated risks. 24

Figure 14. Participants' mean eye fixations compared to roadside pole locations for three video segments. 25

Figure 15. Participants' mean eye fixations compared to a roadside ditch location for a 2.5 sec video segment. 26

Figure 16. Participants' mean eye fixations compared to roadside sign locations. 27

Figure 17. Mean number of eyeblinks and blink durations for 20 video clips and the corresponding mean risk ratings. 28

Figure 18. Mean pupil size and number of saccades for 20 video clips and the corresponding mean risk ratings. 29

Figure 19. The car, computer system, video cameras and screen shots (from the video cameras) from the naturalistic driving testing. 30

Figure 20. The route taken by participants during the naturalistic drive. 31

Figure 21. Mean risk ratings of the 13 test roads from the naturalistic drive and the video-based driving simulator ratings. 34

Figure 22. A participant viewing video footage of their drive during verification testing. 35

Figure 23. The frequency of reporting of specific road features as contributing to participants' risk ratings. 37

Executive Summary

It has long been presumed that drivers' perceptions of risk play an important role in guiding their on-road behaviour. The answer to how accurately drivers perceive the momentary risk of a driving situation, however, is not well known. Previous research into drivers' perceptions of risk has shown that drivers do form judgements about the risk of the road and traffic situations they encounter, but when compared to the objective risk of the situation, the accuracy of those judgements appears to be somewhat variable. The research described in this report was undertaken on behalf of the AA Research Foundation "*to conduct road safety research on the relationships between actual road risks and drivers' perceptions of risk and their driving behaviour*."

This research compared drivers' perceptions of the momentary risk for a range of roads to the objective risk associated with those roads through four complementary research tasks. Each of these research tasks involved participants rating the risk of roads, either while watching videos of the roads, looking at still photographs of the roads or driving on the roads. The different research tasks were used to collect a wide range of measures including: momentary ratings of risk, where participants were looking during and just prior to their ratings, and verbal explanations of the reasons for their ratings.

In the first research task, the TARS driving simulator was used to show 69 participants high-definition videos of rural roads, filmed from the drivers' perspective, while they indicated the momentary levels of risk they were experiencing by moving a risk meter mounted on the steering wheel. The objective levels of risk for the roads were calculated using road protection scores from the KiwiRAP database (part of the International Road Assessment Programme). Subsequently, the participants also provided risk estimates for still photos taken from the videos. In the second research task, another group of 10 participants viewed the videos and photos while their eye movements and fixations were recorded.

The results from these two tasks showed that drivers' perceptions of risk were generally in good agreement with the objective risk, but that certain road situations were perceived as being riskier than the objective risk, and perhaps more importantly, the risk of other situations was significantly under-rated. Horizontal curves and narrow lanes were associated with over-rated risk estimates, while intersections and roadside hazards such as narrow road shoulders, power poles and ditches were significantly under-rated. Analysis of eye movements indicated that drivers did not fixate these objects. Participants' pupil size and eye blinks were

highly correlated with their risk ratings. An analysis of the road design elements at 77 locations in the video revealed five road characteristics that predicted nearly 80% of the variance in drivers' risk perceptions; horizontal curvature, lane and shoulder width, gradient, and the presence of median barriers. There was a small but significant tendency for still photos to be rated as riskier than video, but participants' ratings of risk were highly consistent, both across presentation methods, and within each method.

In the third and fourth research tasks, 14 participants drove the actual roads accompanied by a researcher and at pre-determined locations a tone sounded and the driver gave a risk rating. A video recording made during each participant's drive was later shown to them in the laboratory and the participant was prompted to describe the reasons why they rated the risk in the way that they had. The on-road risk ratings showed very good agreement to the ratings provided for the same road locations in the driving simulator. None of the participants mentioned roadside hazards such as ditches and poles as contributing to their feelings of risk, a finding supporting the results of driving simulator and eye-tracking testing.

Overall, the findings clearly indicate that drivers do perceive and consider risk while driving on New Zealand roads. Although their perception of risk is generally a good match with the objective risk of those roads, there are some specific road features where the risk is under-rated by drivers. These under-rated features include intersections and roadside hazards such as ditches and poles which apparently are not even looked at by drivers. Instead, drivers use curves, hills and road width to judge the risk of a road. The identification of situations with under-rated risks has clear implications for rural highway design in New Zealand as these are the situations where drivers are less likely to take due care and most likely to be involved in a serious crash.

Reading the Risk of New Zealand Roads:
A Comparison of Actual and Perceived Driving Risk

1. Background

Since the earliest days of research into driver behaviour, it has been reported that that drivers modify their behaviour according to the risk they perceive (Fuller, 2005; Gibson & Crooks, 1938; Näätänen & Summala, 1974; Taylor, 1964; Watts & Quimby, 1980; Wilde, 1982). It has even been proposed that this behavioural factor is the most important of the three main factors associated with road crashes: behavioural, vehicular, and environmental (Armsby, Boyle & Wright, 1989). Unfortunately, drivers do not always accurately perceive hazards and risks, and as a result, their behaviour may not be appropriate to the circumstances.

Gibson and Crooks proposed that drivers adjust their speed and lane position according to a perceived "field of safe travel" (Gibson & Crooks, 1938). Gibson and Crooks hypothesised that drivers perceive a safety zone around their car and in their projected path ahead. Implicit in this account is that drivers avoid departures from the safety zone because of the perceived risk of collision and bodily injury with other cars and obstacles.

Developing a similar line of thinking, Taylor (1964) proposed that drivers' level of emotional tension or anxiety resulting from their distance to possible hazards served to govern their speed and lane position. Taylor based his proposal on the measurement of galvanic skin responses (GSR) in two on-road experiments in which 20 participants drove roads of varying levels of difficulty and accident risk. Taylor reported that GSR levels for various road sections were correlated with accident probabilities and inversely related to driver speed during those road sections. Taylor suggested that drivers regulated their own driving by adjusting their speed to maintain their level of anxiety, fear, and tension within acceptable levels.

In a widely-cited series of papers, Wilde (1982, 1988, 2002) elaborated these ideas into what he called the Theory of Risk Homeostasis. Wilde's theory proposed that drivers possess an internal, target level of risk and they will increase or decrease the safety of their driving in order to reduce the difference between their momentary perceived level of situational risk and their target level (Wilde, 1988). Wilde's interpretation of Taylor's findings was that increases in anxiety or arousal (as indicated by GSR) reflected drivers' perception of a level of risk that was higher than that which they personally considered acceptable or safe. The large individual differences in GSR activation noted by Taylor represented the different "set points" for allowable risk maintained by individual drivers.

The notion of risk compensation even seemed to be presaged by the original work of Gibson and Crooks, who wrote "More efficient brakes on an automobile will not in themselves make driving the automobile any safer. Better brakes will reduce the absolute size of the minimum stopping zone, it is true, but the driver soon learns this new zone and, since it is his field zone ratio which remains constant, he allows only the same relative margin between field and zone as before" (1938, p. 458).

The Theory of Risk Homeostasis was not without controversy however. Researchers have been critical of the theory for a variety of reasons, including: incorrect statistical interpretation of aggregated data used to support the theory (Lund & O'Neill, 1986); logical inconsistencies internal to the theory (McKenna, 1987); a lack of evidence that groups act to maintain a constant level of risk (Evans, 1995); and a level of ambiguity in the theory that made it very difficult to generate testable predictions (Elvik, 2004). It has been proposed that these changes in driver behaviour are better thought of as "behavioural adaptation" so as to avoid the temptation to view them as necessarily an outcome of changes in the levels of risk or a conscious decision by drivers in response to their perceptions of risk (Lewis-Evans & Charlton, 2006). For example: "the consensus is now that although drivers tend to adapt their behaviour to improved road and vehicle engineering design, they do not in all circumstances adapt their behaviour such that risk remains constant" (Rothengatter, 2002, p. 251).

Further to this point, drivers appear to focus on factors contributing to the risk of crash occurrence rather than risk factors associated with the severity of a crash. Lund and O'Neil (1986) were the first to suggest that offsetting behaviour (behavioural adaptation) would be more likely to occur for changes that affect crash probability (studded tyres) than changes that affect injury probability (seat belts). They proposed that because changes that reduce the likelihood of a crash also often provide direct and immediate feedback, drivers may be more likely to change their behaviour. In contrast, changes that increase occupant protection usually do not provide direct and immediate feedback to the driver and, therefore, should have no little or no effect on driving behaviour. Their analysis of the existing research was consistent with their prediction; behavioural adaptation is more likely to occur for accident-reducing rather than for injury-reducing measures (Lund and O'Neil, 1986). Sagberg et al. (1997) reached a similar conclusion when their analyses also confirmed that accident-reducing measures like ABS are compensated for by road users' behaviour to a larger extent than injury-reducing measures such as an airbag. Similarly, Harless and Hoffer (2003) concluded that there is stronger support for behavioural adaptation when studded snow tires are adopted or when weather or lighting affect driving conditions (and hence the likelihood of an accident) whereas evidence of risk compensation is weaker or absent altogether for injury-reducing measures.

Similar findings have been reported for a range of safety improvements made in road designs. For example, Assum et al., (1999) hypothesised that introduction of road lighting would have little effect of drivers' behaviour, based on previous research. They found, however, that drivers' speeds increased significantly by 3.6 km/h after road lighting was introduced, and were 5% higher than a control section of highway. The researchers also reported that drivers' concentration was reduced by road lighting (as measured by lane position variability). In a large study of roads in the UK it was found that following road resurfacing, traffic speeds can increase by up to 2.6 km/h (Cooper, Jordan, & Young, 1980). A similar study of road surfaces in Finland reported that resurfacing increases the average traffic speeds, at least when the road is dry (Leden, Hämäläinen, & Manninen, 1998). In New Zealand it has been established that drivers' speeds increase following curve realignments (Wong & Nicholson, 1992) and they drive at higher speeds when road width increases (Burdett & Nicholson, 2010; Charlton & Baas, 2006).

It may be useful at this point for a brief clarification of definitions. A hazard is an aspect of the road environment, or any combination of circumstances on the road that poses a danger to drivers (i.e., increases the likelihood of a crash). Objective risk is the likelihood of a crash associated with the presence of a specific hazard as calculated by analysis of actuarial data from previous crashes. Objective risk is usually expressed in terms of statistics such as probability values or odds ratios. By contrast, subjective risk (perceived risk) is the level of danger associated with a hazard, as perceived by an individual. Subjective risk is rarely quantified in absolute terms, but drivers appear capable of making comparisons between risk levels at different sites and detecting changes at individual sites over time. Subjective risk is thus in the eye of the beholder whereas a hazard is a property of the driving environment and as such there may be true hazards in an environment that individual drivers do not notice, or that they notice but do not consider to be a risk (Armsby, Boyle, & Wright, 1989). This distinction is of significance to the present discussion since it is important that road hazards be noticeable to drivers and perceived as a driving risk. As Gibson and Crooks noted;

"hidden obstacles are dangerous, when they are, because they tend to put the driver's field of safe travel out of correspondence with reality" (1938, p. 471).

In an early study of drivers' perceptions of risk, Pelz and Krupat (1974) showed 60 undergraduate men a 5 minute film of highway driving as seen from the driver's seat and recorded moment-to-moment judgments of danger by means of an "apprehension meter". While watching the film, the participants moved a lever with a scale marked SAFE at one end and UNSAFE at the other according to how safe or unsafe they felt as a driver throughout the film. The participants were divided into three groups based on their driving records: Safe Record, Accidents only, and Violations-or-both groups. The Safe Record group had the highest baseline level of caution between hazards and the longest duration of elevated caution for each hazard; i.e., the participants in this group recognized driving risk sooner and longer. Pelz and Krupat also measured the basal resistance levels (BRLs) and galvanic skin responses (GSRs) of the participants as they watched the film and found positive correlations between the ratings of danger obtained with the apprehension meter and the physiological skin-resistance measures indicating that perceptions of risk were positively related to psychophysiological arousal.

Using pairs of still photos taken moments apart, Benda and Hoyos (1983) asked participants to sort 39 different traffic situations (that showed various road and weather conditions) according to their "hazardousness". Their results showed that experienced drivers were able to construct a ranked order of the hazardousness of driving scenes but less experienced drivers (roughly half the years' of experience) grouped the photos according to the type of hazard and did not differentiate the different level of hazardousness shown. The authors also reported that when similar scenes were shown in motion via film clips an equivalent pattern of results was produced and that the participants in each of the experimental conditions tended to separate "comfortable driving" from all other situations. According to Benda and Hoyos, comfortable driving "means driving under good conditions in which drivers do not need to process too much information… relatively few control activities are required. This kind of driving is obviously regarded as fairly nonhazardous" p. 8). Based on this finding they suggested that drivers' perceptions of hazardousness, or subjective risk, depends on both their amount of experience with various sorts of driving hazards and the information load in the situation, higher information loads leading to higher levels of subjective risk.

Watts and Quimby (1980) asked 60 drivers to make assessments of risks along a 16 mile (25.75 km) route on a rural road and compared the participants' risk ratings to objective risk (calculated from crash data and the participants' speeds). The correlation between the objective risk and the participants' subjective risk was only moderate (Spearman's rho = 0.37) and there were many locations where the risks were underestimated or overestimated. Watts and Quimby suggested that the low levels of perceived risk at some sites may have contributed to the high levels of objective risk, and conversely, there was no crash history at the five locations receiving the highest risk ratings, perhaps as a result of the high levels of perceived risk. In order for this presumed relationship between perceived risk and crash history to be true, however there must be some correspondence between subjective risk and the safety margins adopted by drivers. Watts and Quimby noted that in general the safety margins maintained by drivers correlated well with their subjective risk ratings for those sites, but there was one noteworthy location on the route where this relationship was not maintained. At this location, a left hand curve edged with a 1.5 m hedge that limited the forward visibility, the drivers displayed both high speeds and high risk ratings. Although the drivers did note an increase in subjective risk, they were apparently willing to accept the increase in risk rather than slow down to maintain an adequate safety margin.

Similarly, Kanellaidis and Dimitropoulos (1994) compared drivers' ratings of subjective risk and the objective risk for five curves on a four-lane divided arterial road in Athens. Thirty-four volunteer drivers drove the 3 km section of road in each direction and subjective risk ratings were given verbally at the midpoint of each curve. Measurement of objective risk at the curves was calculated by filming the curves from both directions and rating the road elements according to the German Guide for Traffic Evaluation of Highways. A very good correspondence between the objective and subjective risk values was observed for the curves (Spearman's rho = 0.78). Two of the curves on the route were generally regarded as accident "black spots" (although no crash history data were provided in the report) and the greatest discrepancy between objective risk and subjective risk ratings occurred at one of these curves.

In a follow-on study, Kanellaidis, Zervas, and Karagioules (2000) followed a similar procedure for three different road sections and compared a group of 96 drivers aged 18 to 64 to a group of 40 drivers aged 65 to 75 years. The analysis revealed, once again, that differences between actual risk and perceived risk were associated with increased accident frequency, and that in these cases (where subjective risk is viewed lower than the objective risk) the presence of warning signs becomes most important in maintaining adequate safety margins. The researchers also reported that subjective risk ratings increased with drivers' age, the drivers' familiarity with the roads, and self-assessment of driving skill (the higher the rating of driving skill, the lower the rating of subjective risk).

Groeger and Chapman (1996) showed films of 24 road situations to 64 participants seated at the steering wheel of a partial car. At the conclusion of each scene the participants answered several questions about the level of risk depicted in the scene, how much driving skill was required by a driver in that situation, the amount of control over the danger they would have as a driver in that situation. Detailed analysis of the participants' ratings indicated that drivers responded to three main characteristics of the situation when considering the road scenes: danger, difficulty, and controllability/abnormality (their level of control and what they would normally expect in that situation). Examining how these factors affected drivers of different ages and experience levels, Groeger and Chapman reported that young drivers tended to respond to the danger of the situation more than the difficulty involved in manoeuvring and appeared to consider the situations less dangerous compared to older drivers. Based on this, and other research, Groeger and Chapman pointed out that although drivers attended to these three factors in their judgements of driving situations, it does not necessarily mean that their judgements of subjective risk were accurate. They argued that, in fact, there is compelling evidence that ratings of subjective risk are highly unreliable and prone to distortions associated with the context in which the judgements are made.

2. Research goal and approach

The goal of the project was to investigate the relationship between drivers' perceived levels of risk (that arise from the appearance of roads) and the objective levels of risk associated with those roads. Specifically, the research was undertaken on behalf of the AA Research Foundation "*to conduct road safety research on the relationships between actual road risks and drivers' perceptions of risk and their driving behaviour.*"

Previous research into drivers' perceptions of risk has shown that drivers do form judgements about the risk of the road and traffic situations they encounter. When compared to the objective risk of the situation, the accuracy of those judgements appears to be somewhat variable (Charlton, 2011). Asymmetries between perceived risk and objective risk have been reported in a range of published reports and appear to be a joint result of driver characteristics (e.g., experience) and the visual characteristics of the roadway (e.g., Howarth, 1988; Joshi, Senior, & Smith, 2001; Watts & Quimby, 1980).

Aside from the theoretical interest in the correspondence between drivers' perceptions of risk and the objective risk of various driving situations, there are clear practical reasons for investigating this relationship. Specifically, the published literature suggests that sections of road where drivers' perceived risk is significantly lower than the objective risk (known as risk discordance) present a significant hazard to drivers (Kanellaidis & Dimitropoulos, 1994; Watts & Quimby, 1980).

In order to address the research goal, the following research questions were formulated:

Question 1 - *What levels of subjective risk are experienced on hazardous New Zealand roads?*

Question 2 - *What road features do drivers use to judge driving risk?*

Question 3 - *What hazardous road situations are under-recognised by New Zealand drivers (i.e., show the greatest dissociation between objective and subjective risk)?*

[Note: A fourth question, "*What countermeasures can be used to convey a more accurate perception of risk?* was addressed by Beca Ltd in a parallel research activity.]

In order to address these three research questions a range of complementary research methods, including both laboratory and on-road methodologies, were employed as shown in Figure 1. As can be seen in the figure, the six research tasks (four of them test methodologies) represented complementary approaches to obtaining measures of subjective risk and were selected to provide a high degree of cross-validation of the results.

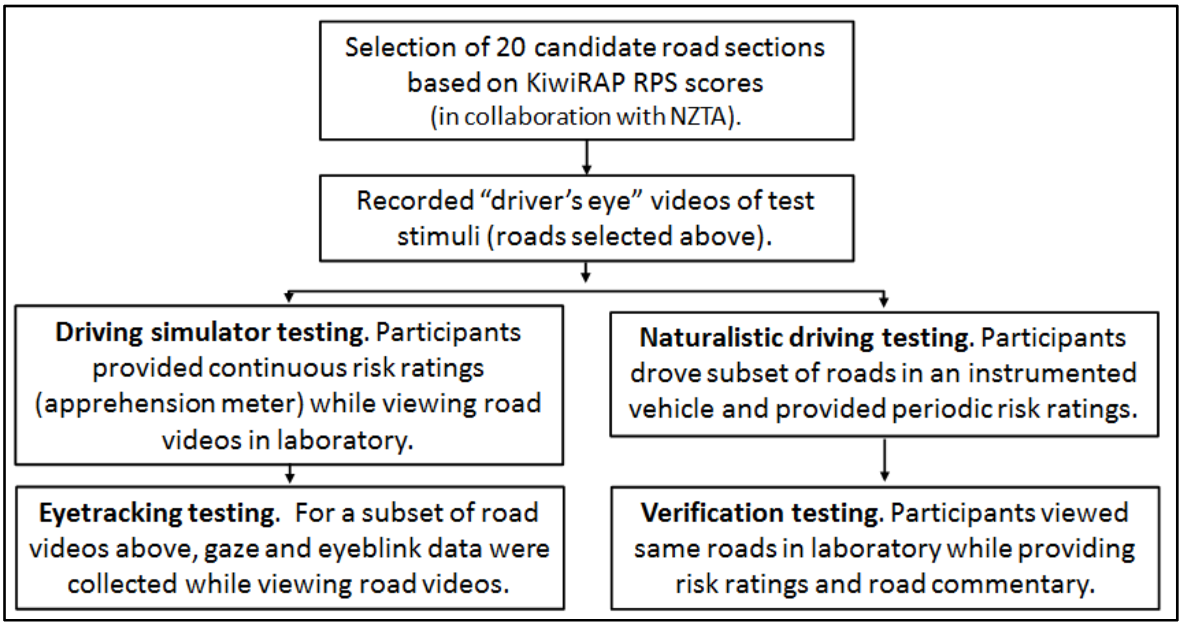

Figure 1. An overview of the six research tasks used to answer the research questions.

For example, explicit measures of subjective risk (risk ratings) were obtained from the driving simulator testing and compared to the objective risk of a range of different driving situations. The risk ratings from this task were collected from both high definition videos and still photos taken from a subset of the videos, allowing a comparison of risk perceptions for these two methods of stimulus presentation. In addition, each of these methods had contained within them some overlap and repetition allowing analysis of the reliability of the participants' ratings. Further, the explicit ratings were compared to implicit measures of risk such as the gaze, pupil dilation, and blink rates from the eye-tracking testing which used the same video stimuli.

Finally, the on-road naturalistic driving task collected risk ratings from the same points of interest contained in the videos. These risk ratings, and the participants' reasons for their risk ratings provided during the verification task, provided another point of comparison and cross-validation, as well as an explicit narrative to aid in interpretation of the data collected.

3. Driving simulator testing

3.1. Selection of study roads

A representative set of rural roads from the state highway system were selected (in conjunction with NZTA representatives) such that they were representative of a range of objective risk. The measure of objective risk used was the Road Protection Score from KiwiRAP (the New Zealand Road Assessment Programme: http://www.kiwirap.co.nz/index.html). Road Protection Scores (RPS) are calculated for each 100m section of state highway based on a formula incorporating the risk of three primary crash types (run-off road, head-on and intersection), road design features most likely to influence the occurrence of crashes (e.g., road type, lane and shoulder width, sight distance, overtaking provision, roadside features) and traffic volume. Levels of collective risk, as determined by crash histories recorded in the NZTA's Crash Analysis System database, were also used during the selection process. A total of 36 sites were identified from state highways SH1, SH1B, SH3, SH31 and SH39. All of the sites of interest were located within the Waikato region.

3.2. Recording test stimuli

High-definition video (HD resolution, 60 Hz frame rate) of the selected roads were collected from a specially-configured video capture vehicle driven at the posted speed limits in a safe (i.e., non-aggressive) driving style by an experienced driver. The camera was placed to approximate a driver's perspective while still ensuring the vehicle could be operated in a safe manner. The video collection was repeated several times at each location to ensure that the visibility was clear, weather conditions were comparable, and the presence of other vehicles was minimised (so that the level of risk depicted in the videos arose primarily from the road features rather than momentary conditions or the behaviour of other road users). Each of the videos, with accompanying car and road sounds, was edited into a series of 45 sec "clips" or test stimuli and joined into a 1,690 sec (28 min 10 sec) test video in which each clip was separated from adjacent clips by a 2 sec interval which dissolved from the clip to black and then to the next scene. The resulting video contained a total of approximately 45 km of driving across the 36 locations. Three versions of the testing video were created containing the 36 video clips in one of three random orders, each of them beginning with the same clip (used as a control or "warm-up" to start the video). A 188 sec (3 min 8 sec) training video containing four clips not used in the test stimuli video was created and used during participant familiarisation.

Four clips were cut so that they overlapped in two of the sections of road depicted (two clips had 35 sec of overlap, and two other clips had 15 sec of overlap). These clips were used as a reliability/consistency check on the participants' ratings and when they were placed in the videos so that they were never adjacent to one another.

In addition to the test videos, a set of still photos were created (1920 by 1200 pixels) for use as recognition test/risk rating stimuli. Twelve still photos were taken from frames of the test videos, and 14 other photos taken from the raw video (not contained in the 36 test clips or at locations near them). Two of the "new" photos were designated as practice stimuli, and the remaining 24 photos were put into three random orders for presentation to participants. One of the photos taken from the test video was used as a reliability check and was repeated three times in each presentation order (in nonadjacent locations).

3.3. Method

3.3.1. Participants. Seventy-four participants with full (unrestricted) NZ driving licences were recruited to take part in the study, 69 of whom completed the full test protocol (five participants did not complete the testing due to equipment difficulties, eyestrain, or other reasons). The 69 participants (31 males; 38 females) had an average age of 33.7 years (range 17- 69 years), and reported an average of 15.7 years since receiving their full NZ licence (range 1.6 – 50 years). 47.8% of the participants reported that they had no crashes and 52.2% reported being involved in one or more crash. 75% reported that they had no driving infringements; 21.7% reported 1 or more. The self-reported ethnicity of the sample was 71% European, 10% Maori, and 18% Other. Ethical approval for the recruitment and test protocols was received from the School of Psychology Research Ethics Committee at the University of Waikato. Each of the participants received a $20 gift voucher for this stage of the testing.

3.3.2. Apparatus. The TARS driving simulator consisting of a complete automobile (BMW 314i) positioned in front of a projection surface on which the test stimuli were projected 2.64 m wide by 2.10 m high (at a resolution of 1920 by 1200 pixels). Details of the TARS simulator have been described elsewhere (Charlton & Starkey, 2011).

3.3.3. Procedure. Following informed consent and completion of a brief demographic questionnaire the participants were seated in the simulator and shown the training video. During the training video the participants were asked to engage in a steering task in which they were instructed to keep a circle lined up on the centreline of the road by means of the steering wheel in front of them (the same secondary steering task as used by Charlton, 2006). The instructions to the participants were as follows:

> *We are going to show you a video of several local roads. We want you to imagine yourself the driver actually driving on these roads. While you watch we want you to move the steering wheel as you would on the actual road. During the practice video we will put a yellow dot on the screen so you can see how the car responds to your steering.*

Participants were also instructed to provide moment-to-moment judgments of driving risk by means of an "apprehension meter" (an analogue to Pelz & Krupat, 1974). The apprehension meter consisted of a thumbwheel mounted on the right side of the steering wheel (see Figure 2). The thumbwheel controlled a pointer which moved along an on-screen scale anchored with the words "Safe" at the bottom and "Unsafe" at the top with nine calibration lines between (see Figure 3). The instructions to the participants for this part of the task were as follows:

> *The main thing we want you to do is to report how safe or unsafe you feel as the driver at all times. On the right of the steering wheel is a thumb wheel and by moving this wheel you can move the pointer on a meter that will be shown on-screen in front of you.*

If you felt completely at ease-if you were at rest or parked and could completely take your mind off driving-you would keep the pointer at the SAFE end of the meter. But, if you felt extremely threatened, very unsafe, or in immediate danger of being involved in a serious accident or mishap you would move the pointer all the way to the UNSAFE mark at the top. Keep the meter pointed to the position along the scale that best expresses your feelings of risk throughout the video. The screen will go dark briefly In between the different roads shown in the video, you should move the meter all the way down to the SAFE end during these brief pauses

Figure 2. Thumbwheel used by participants to provide risk ratings.

Figure 3. A scene from the test video showing the on-screen risk scale.

The participants were shown the 28 min test video (one of the three orders) during which they provided moment-to-moment risk ratings for the 36 video clips. At the end of the video the participants were invited to take a short (2-3 min) rest break and get out of the car to stretch. The participants were then given instructions about the self-paced recognition and risk rating task for the still photos:

> *Now we are going to show you some photographs, some of them are from the video you just watched, and the others have been taken from different roads at random. For each photograph we want you to tell us whether you recall seeing the road in the video by using the indicator lever on the left of the steering wheel. Then for each photo tell us how risky it looks by using the risk meter. Once you have set your choices, move the headlight lever to record your answers. We will begin with two practice photos.*

An example of one of the stimuli is shown in Figure 4. After providing ratings for the 28 photos (two practice photos, 24 test photos, and two reliability check photos) the testing was concluded and the participants were thanked for their participation and given a $20 gift voucher.

Figure 4. A screen from the still photo recognition/risk rating task.

3.4. Results

Participants' risk ratings during the videos ranged the entire risk rating scale (where "safe" = 1 and "unsafe" = 10); the point in the video with the highest risk rating had a mean of 6.68

(SD = 2.60) and the point with the lowest risk rating had a mean of 1.93 (SD = 0.91). Risk ratings for the 24 photos showed a very similar pattern, with the mean of the participants' ratings ranging from a high of 5.96 (SD = 2.19) to a low of 2.06 (SD = 1.18). No significant differences in risk ratings were identified as resulting from participant gender or age.

For subsequent analysis of the video clips, specific points of interest were identified. 35 (1 per clip) of these were based on the presence of particular hazards (e.g., poles, ditches, vertical or horizontal curves) and another 22 points were identified from the participants' risk ratings (where ratings were high or large changes in ratings occurred). Points of interest that included other road users such as vehicles parked at the roadside, oncoming trucks, and pedestrians were excluded. In addition there were 20 locations depicted in the videos that were selected on the basis of their high RPS (objective risk), yielding a total of 77 locations of interest/points in the video for the analysis of participants' risk ratings. The risk ratings were analysed in three ways: comparison of the ratings of perceived risk with the objective risk, comparison of the ratings of the video with the still photos, and assessment of the consistency of the participants' ratings by comparing the repeated video clips and photos.

3.4.1. Comparison of perceived and objective risk. The first step in comparing the participants' ratings of risk with the objective risk was to rank the 77 road locations of interest according to their RPS score. The risk ratings for the 10 locations with the lowest RPS scores were then averaged and compared to the mean risk rating for the 10 locations with the highest RPS scores. Because of the way RPS scores are calculated from objective risk components, the 10 locations with high RPS scores were all intersections. In order to include other road features in the analysis, the risk ratings for locations with the highest RPS scores excluding intersection were then identified. Finally, the 10 locations around the median RPS score (excluding intersections) were identified and mean risk ratings calculated. Figure 5 shows the results of this step of the analysis, the mean risk ratings associated with each of the four categories of objective risk. A within-subjects analysis of variance revealed a significant difference in risk ratings across the four objective risk categories; $F(3, 204) = 214.78$, $p < .001$, $\eta_p^2 = .760$. As can be seen in the figure, there was generally good correspondence between the participants' ratings of risk and the objective risk, with the notable exception of the "High" RPS category (containing intersections) which were rated approximately the same as the Median risk category locations.

Another view of this relationship can be seen in Figure 6. In the left panel of the figure are the mean risk ratings plotted against their corresponding RPS scores for all 77 locations of interest. The correlation between risk ratings and RPS scores overall is rather low; Pearson's $r(68) = .081$, $p = .486$. In the right panel of the figure are the mean risk ratings and RPS scores for the 57 locations of interest excluding the 20 intersections. The correlation here is strong and positive; Pearson's $r(68) = .771$, $p < 001$. In other words, the participant's risk ratings agreed quite well with the objective risk scores, with the exception of the ratings of intersections, which although having high risk RPS scores, were not rated as high risk by the participants.

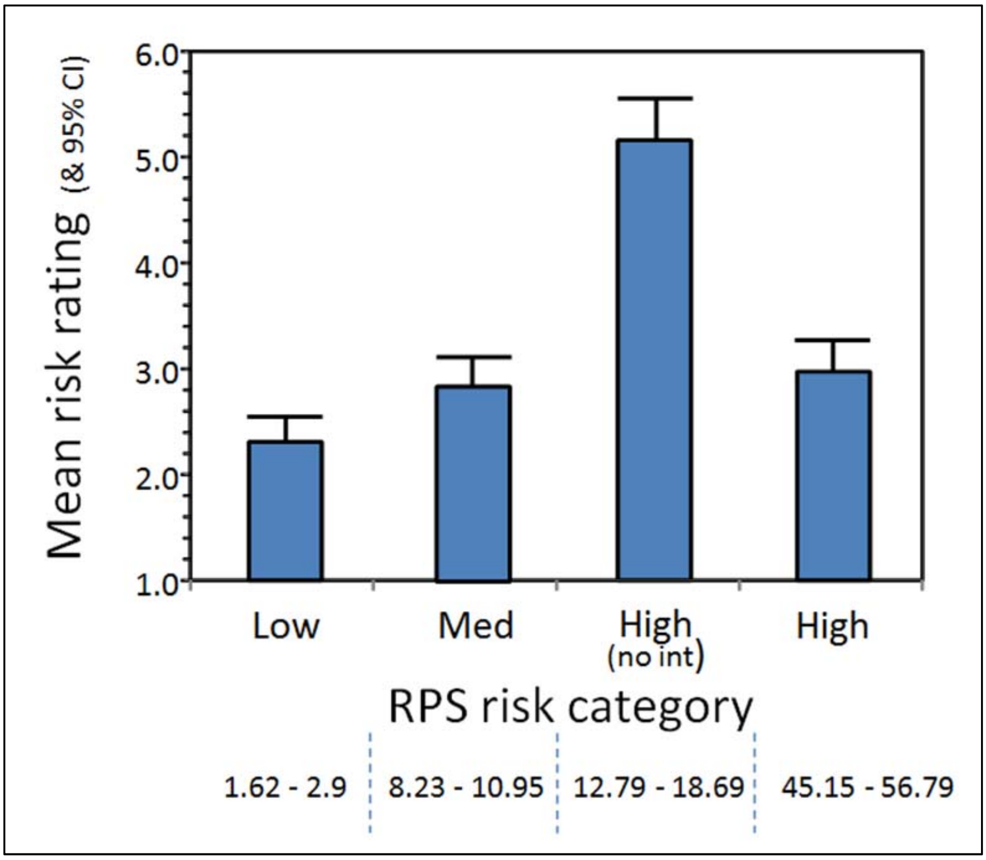

Figure 5. Mean risk ratings for video scenes from four objective risk categories.
The "High (no int)" category represents the 10 highest risk locations excluding intersections.
Lines show 95% confidence intervals associated with the means.

The right panel of the figure also shows some locations where the risk ratings were above the 95% confidence intervals, and others that fell well below the confidence intervals. The locations lying above the interval represent situations that the participants perceived as riskier than the objective risk in the situation. The left panel of Figure 7 identifies the situations giving rise to these over-rated risks; all instances of narrow lane width, horizontal curves, and wire rope barriers are indicated, and as can be seen, most of them fall above the upper confidence interval. The right panel of the figure shows all instances of narrow shoulders, ditches, poles in close proximity to the road, and these constitute the majority of the situations that were under-rated as risks by the participants. Other under-rated risks included banks on the roadside, wide (dual) centre lines, and one location with a grass median (but no barrier).

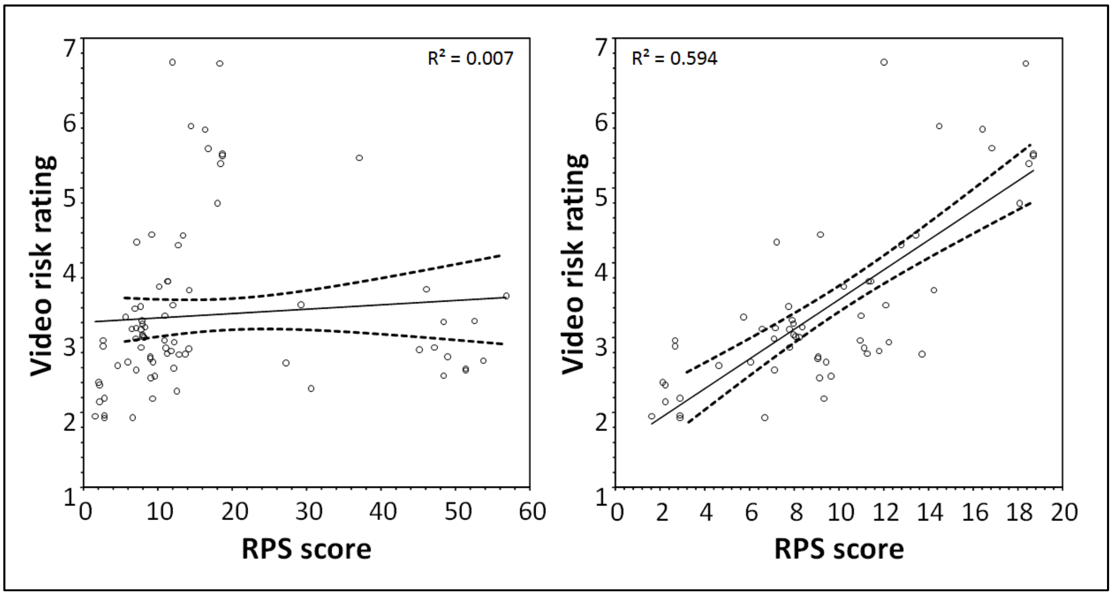

Figure 6. Mean risk ratings and RPS scores for video scenes including (left panel) and excluding the 20 intersections (right panel). Dashed lines show 95% confidence intervals.

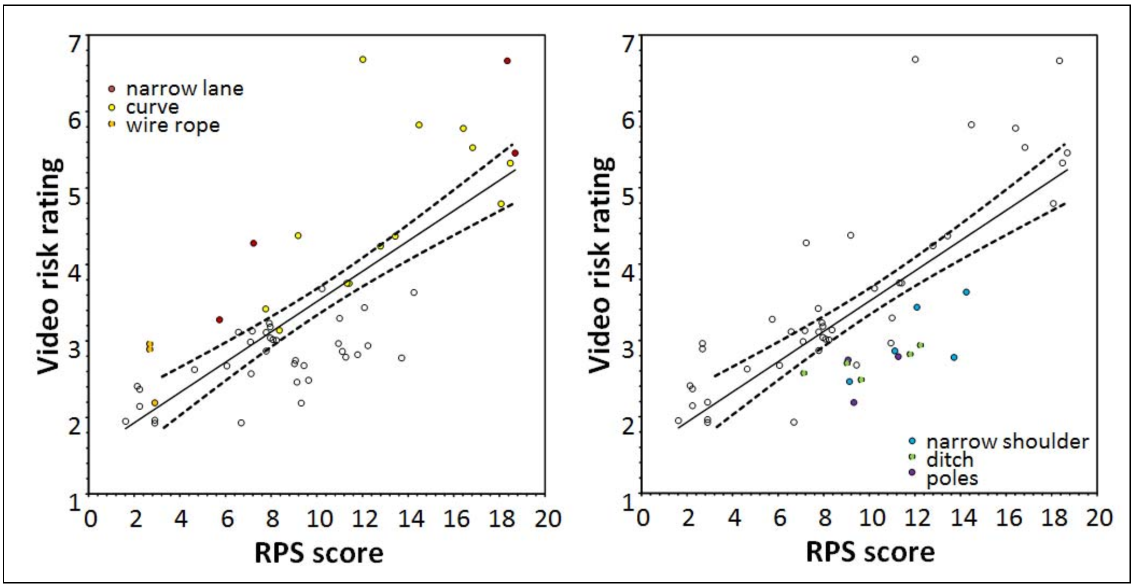

Figure 7. Mean risk ratings and RPS scores for video scenes indicating the types of road situations that were over- (left panel) and under-rated (right panel) by the participants. Dashed lines show 95% confidence intervals.

A multiple regression analysis predicting the mean risk ratings for each of the 77 locations from the 13 components of the RPS scores at those locations revealed that the single largest predictor of the participants' ratings was the Horizontal Alignment Score; Adj R^2 = .532, F (1,73) = 85.00, p < .001. The best combination of RPS components predicting the risk ratings was Horizontal Alignment Score, Lane Width, Shoulder Width, Terrain (gradient), and Right-hand Roadside Risk; Adj R^2 = .785, F (1,69) = 54.97, p < .001. These 5 RPS components together accounted for nearly 80% of the variance in the participants' risk ratings, none of the other component measures were significant predictors of the risk ratings. (It should be noted that in this context the RPS component Right-hand Roadside Risk was a measure of whether or not the opposing lanes were separated by a physical barrier.)

3.4.2. Comparison of video and still photo risk ratings. Figure 8 shows the risk ratings from the 12 photos presented to participants in the second stage of the driving simulator testing compared to the risk ratings for the corresponding locations of the video. As can be seen in the left panel of the figure, the correspondence is very high, statistical analysis indicated a strong positive correlation; Pearson's r(68) = .926, p < .001. Two locations fell above the 95% confidence interval indicating that they appeared riskier in the photos than they did during the video and, as shown on the right of the figure, both locations were bridges.

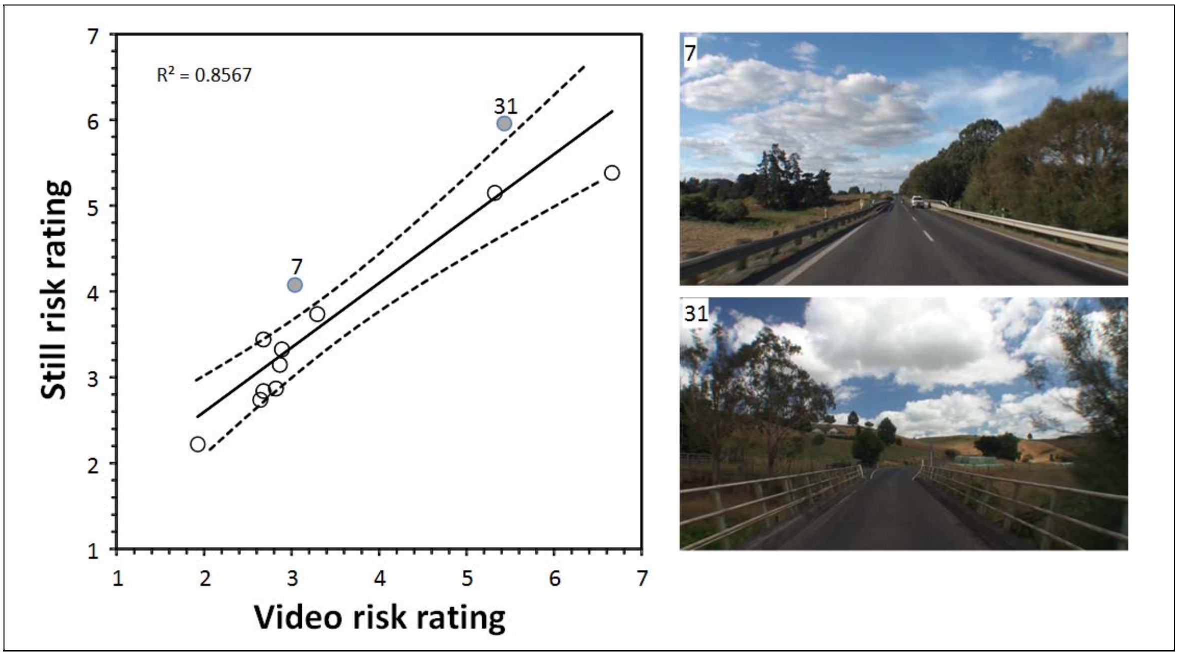

Figure 8. Mean risk ratings for 12 still photos and corresponding video locations of the roads over-rated in the still photos. Dashed lines show 95% confidence intervals.

Overall, there was a slight tendency for the participants to rate the 12 still photos as being riskier than the corresponding locations in the video ($M = 3.74$, $SD = 1.18$; $M = 3.52$, $SD = 1.45$; for photos and video respectively). A repeated measures analysis of variance indicated that this difference in media type was statistically reliable; $F\ (1, 67) = 9.48$, $p = .003$, $\eta_p^2 = .124$. There was also, however, a statistically significant interaction between media type and the individual locations, indicating that the correspondence between the ratings given the stills compared to the video differed somewhat from location to location; $F\ (11, 737) = 11.37$, $p < .001$, $\eta_p^2 = .145$. Analysis of the individual locations revealed that this was restricted to three of the 12 locations; a location with narrow lanes and a wooden edge rail (that was rated riskier in the video), and a bridge (shown as location 7 in Figure 8) and a vertical crest (both rated riskier in the photos than in the video). Comparing the participants' ratings of 10 of the photos to the RPS scores for those locations (RPS scores were unavailable for two locations) produced a strong positive correlation, Pearson's $r(68) = .847$, p = .002; only slightly lower than the correlation between video ratings and RPS scores for the same locations; $r(68) = .886$, $p < .001$.

Finally, the participants correctly recognised 70.41% of the 12 photos as coming from the video, ranging from a high of 92.75% to a low of 44.94% ($SD = 13.45$), the mean increasing to 71.01% if the repeated reliability check photos were included. The participants did not perform as well at rejecting the 12 photos presented that were not from the video, answering a correct "no" response to 49.15% of the photos (SD = 21.34), although this is presumably because these photos depicted the same roads shown in the videos, albeit at different locations. An analysis of the relationship between the risk ratings and the percent of photos correctly recognised failed to indicate any reliable correlation; Pearson's $r(13) = -.088$, $p = .764$, suggesting that the perceived risk of the road was not related to how well it could be recalled.

3.4.3. Consistency of the risk ratings. The consistency of the participants' ratings was examined by comparing the ratings at the two locations that were depicted in overlapping video clips, as well as the ratings obtained for the photo that was presented three times during the second stage of simulator testing. Figure 9 shows the mean of the repeated risk ratings for 30 sec of overlap between two video clips showing the same location. As can be seen, the mean ratings are very consistent, showing nearly identical reactions to the road environment on successive viewings, even though the locations were presented at different points in the two clips, and the clips themselves came at different points in the videos. A consistency check for the ratings of the midpoint of the video shown in the figure across all 69 participants produced a highly significant Cronbach's Alpha of .904. The second section of road presented in two other overlapping video clips displayed a nearly identical pattern, with a Cronbach's Alpha of .888 indicating very high consistency between the successive ratings.

In the still photos presented during the second stage of the simulator testing, the reliability check photo received nearly identical ratings each time it was presented. The means of the participants' risk ratings and the photo are presented in Figure 10. A Cronbach's Alpha indicated high consistency in the ratings across the three viewings (= .925).

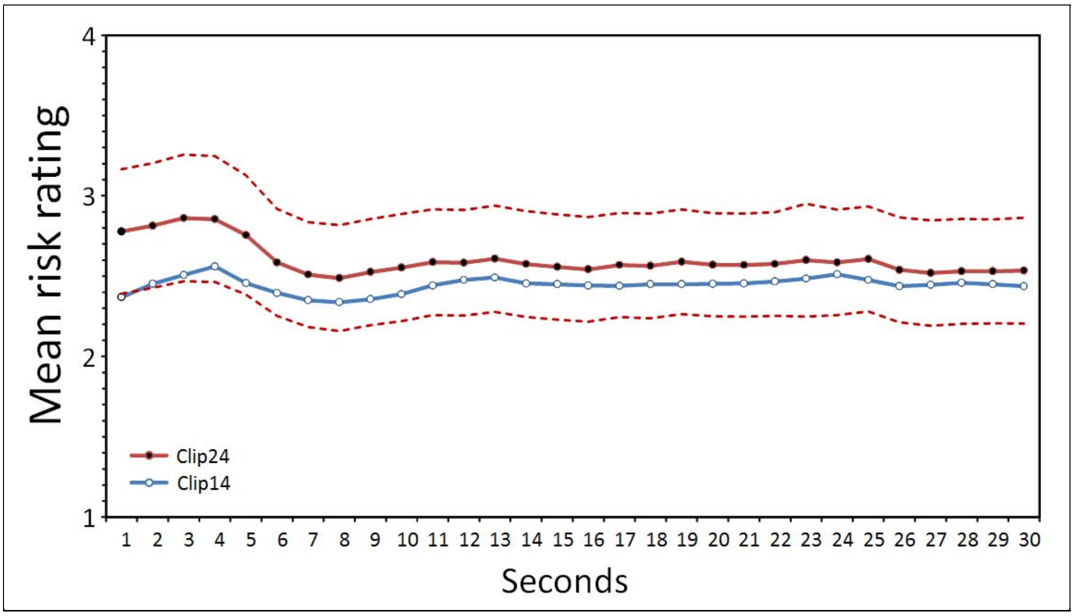

Figure 9. Continuous risk ratings for two viewings of the same section of road. Dashed lines represent 95% confidence intervals for Clip 24.

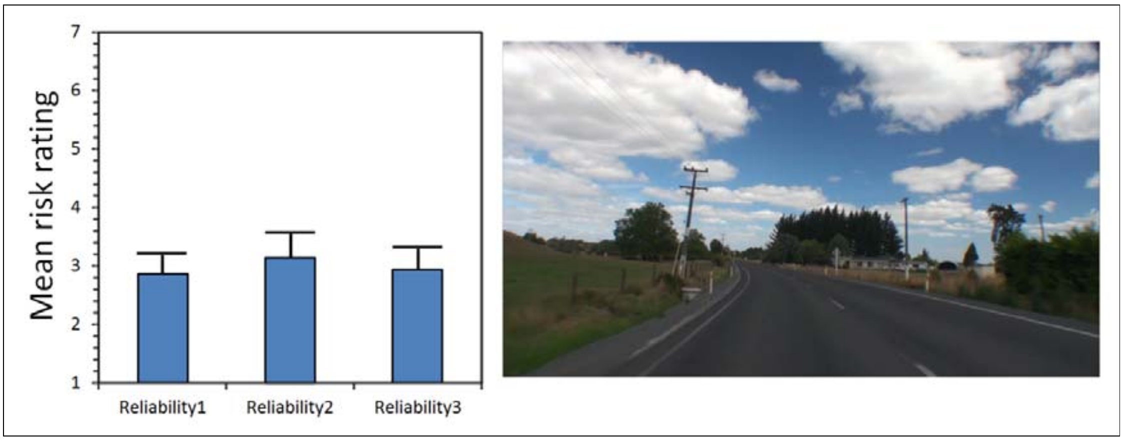

Figure 10. Repeated risk ratings for three viewings of the same photo. Lines represent 95% confidence intervals for the means.

4. Eye-tracking testing

A second sample of participants were recruited to have their eye movements recorded by means of the TARS eye-tracking equipment. The purpose of this was to allow comparison between the explicit risk ratings and implicit measures of risk such as saccades, pupil dilation, and blink rates. The eye tracking data provided information about where participants looked when viewing the videos and providing risk ratings, and which features of the video attracted their gaze.

4.1. Method

4.1.1. Participants. A total of 10 participants were recruited to take part in the eye-tracking testing. Ethical approval for the recruitment and test protocols was received from the School of Psychology Research Ethics Committee at the University of Waikato. Each of the participants received a $20 gift voucher as a token of appreciation. The 10 participants (4 males; 6 females) had an average age of 24.4 years (range 19- 30 years), and reported an average of 8.4 years since receiving their full NZ Licence (range 4 – 13 years). Four of the participants reported that they had no crashes and six reported being involved in one or more crash. Only one of the nine drivers reported that they had received a driving infringement in the past year. The self-reported ethnicity of the sample was 7 European, 1% Maori, and 2% Other.

4.1.2. Apparatus. An SR Research EyeLink II eyetracker (http://www.sr-research.com/eyelinkII.html) set for monocular recording (participant's preferred eye) at a sample rate of 250 Hz was used to monitor eye movements whilst participants viewed a flat-panel display screen (93cm x 52 cm, 1920 x 1080 pixels) from a distance of 90cm (Field of view = 55° horizontal x 32° vertical) (see Figure 11). The tracker recorded (x, y) eye position data, blinks and pupil size. The position data was smoothed with a low-pass filter (-3dB at 32 Hz) and x and y eye velocity (Vx, Vy) was calculated by applying a low-pass differentiator (-3dB at 32 Hz) to the original (x, y) position data. Saccades were detected in the velocity traces using a robust saccade-detection algorithm (Liston, Krukowski & Stone, 2013). The stimulus presentation software and risk indicator were the same as used in the driving simulator testing. As previously described, participants used a thumb wheel attached to a steering wheel to provide continuous (1 Hz) risk ratings during the video clip. The risk scale was superimposed over the right hand side of the video display (15° to the right of the centre of the road) and provided visual feedback to the participants as to their risk rating. Eye position data was analysed for points of interest in the central or left region of the screen only, to avoid the saccades or fixations towards the on-screen risk rating scale.

To minimise discomfort (the eye-tracking equipment becomes uncomfortable if worn for extended periods), a subset of 20 of the video clips used in the driving simulator testing was selected for the eye-tracking testing. The clips were selected to provide maximum overlap with the roads driven in the naturalistic testing and to provide a range of different road types and hazards. These clips, plus the same initial clip used for practice, resulted in a test video 984 sec (16 min 24 sec) in length. The same 3 min 8 sec training video containing 4 clips used in simulator testing was used for instructing the eye-tracking participants. The same set of 24 photos (and 2 practice photos) used for the driving simulator testing were used for the eye-tracking testing.

Figure 11. The TARS eye-tracking equipment recording a participant's eye movements while viewing a road scene.

4.1.3. Procedure. The same informed consent, participant instructions, and training procedure used in the driving simulator testing was used for the eye-tracking testing. As with the driving simulator testing, the participants rated the perceived risk of the road using a risk wheel unit attached to a steering wheel in front of them, while watching the test video (see Figure 11). When the testing was concluded the participants were thanked and given a $20 gift voucher.

4.1.4. Analysis. The eye-tracker data were analysed using custom software (MatLab R2013b, Mathworks) and the blink and saccade totals for each participant and movie clip were extracted. The blink lengths (msec) were also calculated from the blink onset and offset times. Four analyses of the eye movement data were conducted: comparison of the eye tracking risk ratings to the risk ratings from the driving simulator testing; eye-tracking and fixations for under-rated risks; number and duration of eyeblinks in relation to risk ratings; and comparison of pupil size and saccades to risk ratings.

4.2. Results

4.2.1. Comparison of risk ratings. To confirm that the participants in the eye tracking experiment were perceiving the same level of risk as those in the driving simulator testing, we compared the average rating for each location in the test video against the average rating for the corresponding locations in the driving simulator testing. The risk rating data from one of the participants was not recorded because of a computer software problem and so only 9 participants' data were included in this comparison. As shown in Figure 12, the correspondence between the risk ratings obtained for the two types of testing was high and statistical analysis indicated a strong positive correlation; Pearson's $r(19) = .778$, $p < .001$. The mean risk ratings for the two test protocols were nearly identical ($M = 3.41$ for the driving simulator, $M = 3.49$ for the eye-tracking test), although the variability was higher in the simulator ($SD = 1.15$, $SD = 0.76$, for the simulator and eye-tracking test respectively).

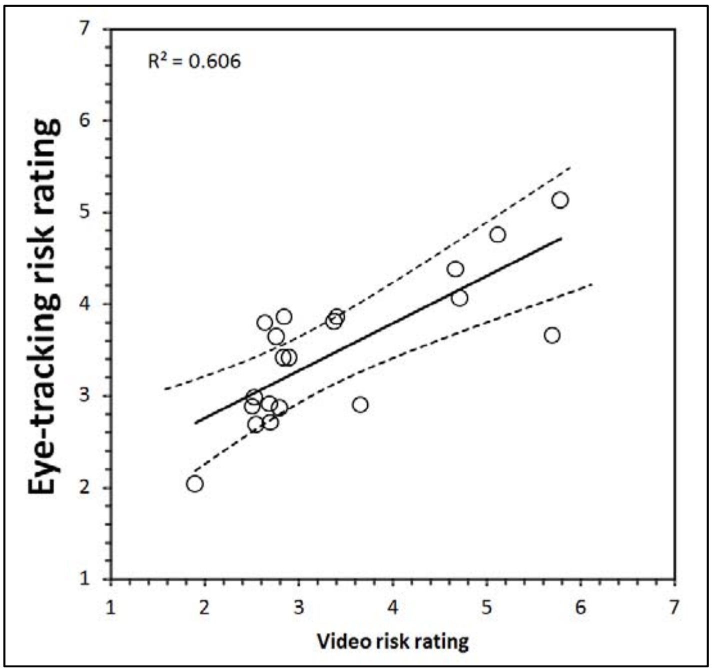

Figure 12. Mean risk ratings for 20 locations in the eye-tracking video and corresponding locations in the driving simulator test video. Dashed lines show 95% confidence intervals.

4.2.2. Eye-tracking and fixations for under-rated risks. The driving simulator results showed that participants rated roadside hazards such as power poles and ditches as low risk even though the objective risk (RPS) is high. A useful question to ask is whether or not drivers noticed these hazards or attended to them. To answer this question, the participants' eye-tracking and fixations were analysed to determine if the participants looked in the direction of the poles. If they did not look at them, then it may suggest that they did not think them sufficiently important to fixate or to pursue with their eyes.

To address this question, a video clip containing a straight piece of road with a series of power poles close to the edge of the road was selected for detailed analysis. In the video clips, these move from the centre of the screen to the left edge over a period of about 2.5 sec as the driver approaches and passes a pole (see Figure 13). For each pole the X location at the start and end of a series of 2.5 sec segments from the video was located.

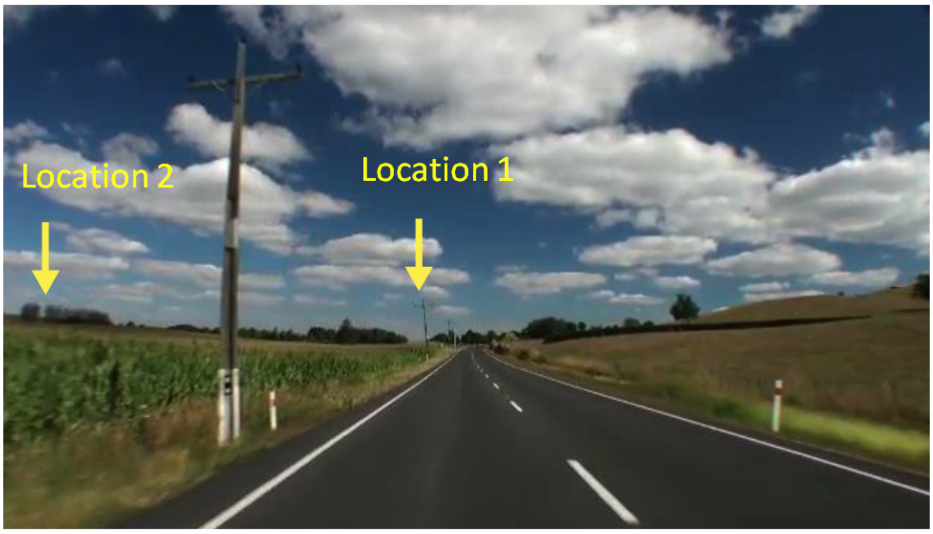

Figure 13. Screen positions of a roadside pole used in the analysis of participants' eye-tracking and fixations for under-rated risks.

An example of pole locations for three of these video segments is shown as a red line in the three panels of Figure 14. For these three example segments, the position of the poles were tracked beginning at image pixel location 866 and ended up at a point 32 pixels from the left edge of the screen. (For purposes of this analysis, all positions to the right of the centre of the screen were set to 0 in order to include only those eye movements on the same side of the road as the pole). The mean X eye location across all 10 participants over these 2.5 sec segments of the video clip were calculated from a series of 4 msecs time samples, along with the 95% confidence intervals for the mean as shown in Figure 14. As can be seen in the figure, except for the very start of the segment, where the pole was close to the centre of the road in the movie image, on average the participants' gaze did not include the position of the pole.

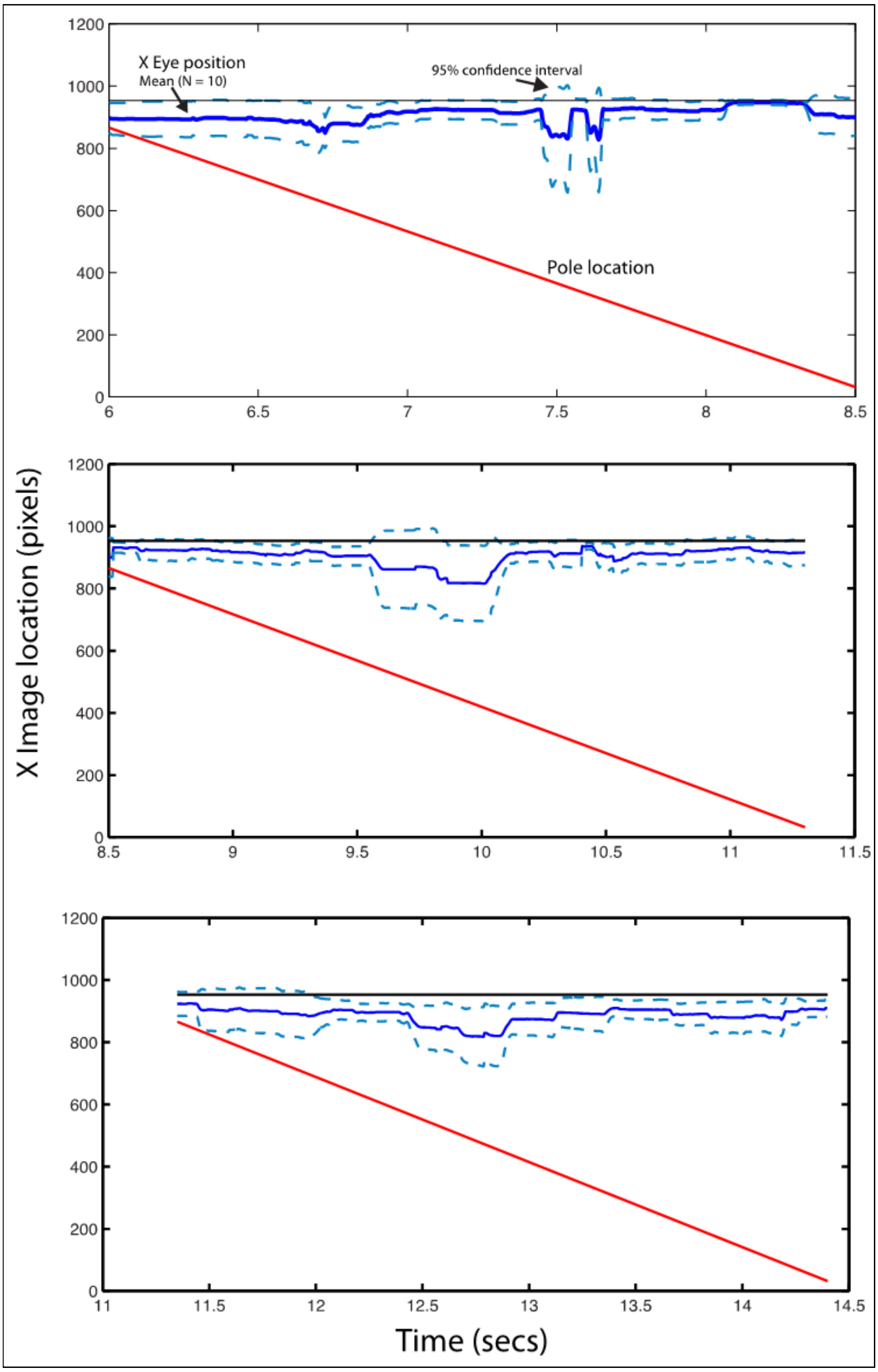

Figure 14. Participants' mean eye fixations (blue lines) compared to roadside pole locations (red lines) for three video segments. Dashed lines indicate 95% confidence intervals for participants' eye fixations.

Roadside ditches were also underrated (relative to RPS scores) by the participants. We carried out a similar analysis to that used for the power poles and examined participants' average fixation locations relative to the position of a ditch to determine whether or not they fixated roadside ditches. The location of a ditch in the video requires both an X and Y position to define and it is not possible to simply compare the eye fixations relative to a line. For this analysis the average (x, y) location across a 5 sec segment of a video clip containing a prominent ditch at the side of a straight section of road was calculated

Figure 15 shows the 10 participants' mean eye fixations superimposed on an image from the middle of the video segment distribution relative to the ditch location (indicated with the dashed oval in the figure). As can be seen in the figure, none of the participants made eye fixations in the location of the roadside ditch.

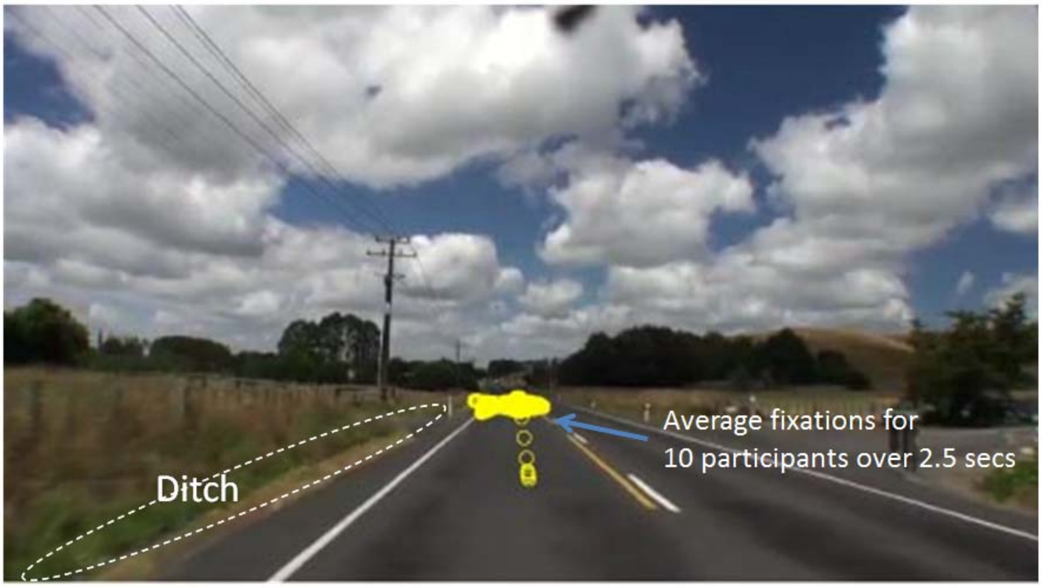

Figure 15. Participants' mean eye fixations (yellow circles) compared to a roadside ditch location (dashed lines) for a 2.5 sec video segment.

To investigate whether the above results could be a consequence of the participants simply not fixating any roadside objects, a similar 2.5 section of straight road containing a roadside object other than a pole or ditch was selected from the video. The roadside object was a PW17 advance curve warning with a supplementary speed advisory (see Figure 16). Over the course of the 2.5 sec video the sign moved from location 853 to 613 pixels from the left of the screen. The mean X eye position for the participants relative to the position of the roadside sign across the 2.5 sec video is shown in the figure, and as can be seen, a substantial number of eye fixations did occur to the roadside area containing the sign. (In this example the car was travelling slower and so the red line indicating the sign position is less steep than in the roadside pole examples). From this example it is apparent that the participants' functional field of view did include roadside objects, however, the roadside poles and ditch did not attract fixations to the degree that the example curve advisory sign did.

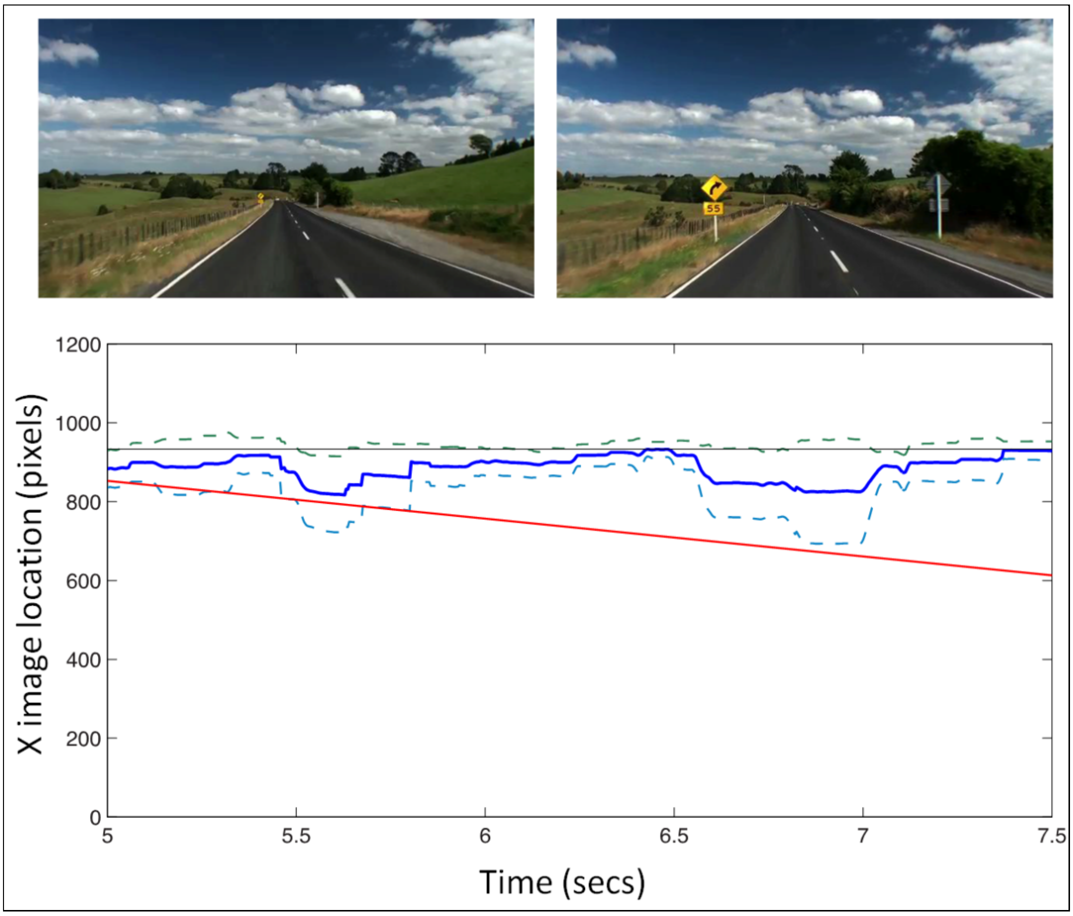

Figure 16. Participants' mean eye fixations (blue line) compared to roadside sign locations (red line). Dashed lines indicate 95% confidence intervals for participants' eye fixations.

4.2.3. Analysis of blinks. Previous research has shown that blink rate is a good indicator of attentional demands (cognitive workload) and stress (Brookhuis & de Waard, 2010). Given this, it was of interest to explore the relationship between blinks and subjective risk in the current study. First, the total number of blinks that occurred (over the total length of the clip) and the duration of each blink were calculated for each location in the test video. Linear regression analyses were then conducted for each of these variables against the average risk rating for the clip (over the total length) to see if there was any relationship between the number and duration of blinks and the risk ratings. The results of these analyses are shown in Figure 17. As shown in Figure 17, both the number and length of the blinks decreased as the risk rating increased. Statistical analysis indicated that both of these relationships were significant, Pearson's *r*(19) = -.534, *p* = .015 for number of blinks and *r*(19) = .-553 for blink length (*p* = .011).

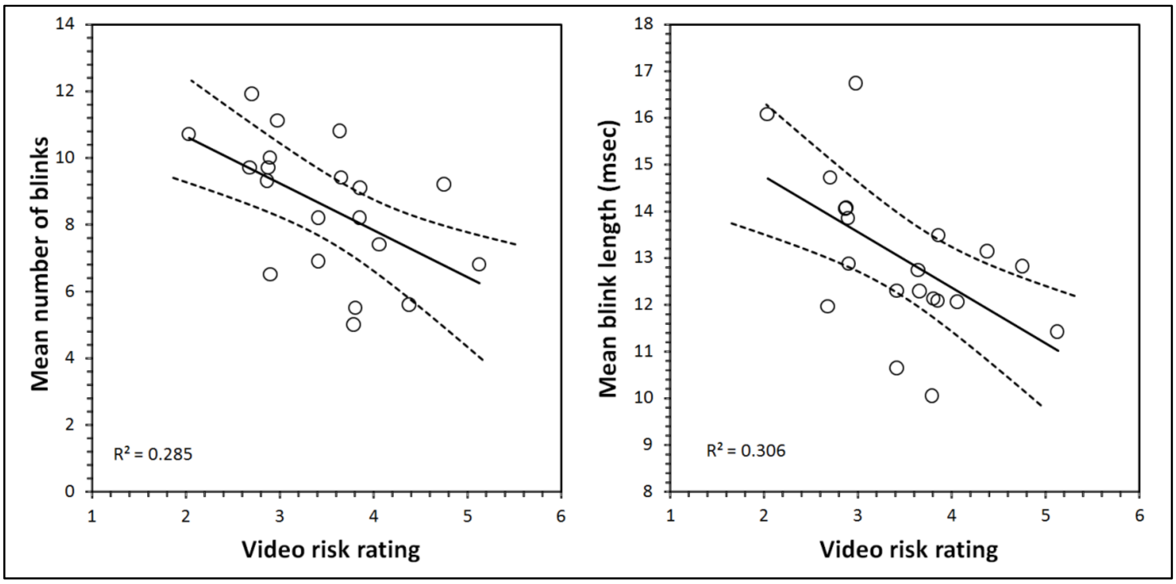

Figure 17. Mean number of eyeblinks (left) and blink durations (right) for 20 video clips and the corresponding mean risk ratings. Dashed lines indicate 95% confidence intervals.

In order to examine whether or not eyeblinks corresponded to perceptions of risk at specific locations, as opposed to averaging across the 45 sec duration of each video clip, each video clip was divided into 'Low risk' (where the risk rating was <= 5) and 'High risk' (> 5) segments. For each segment the total blink length (i.e., total eye closure time) was used to calculate the blink density (total eye closure time/ length of segment) for each of the high and low risk locations in the video. This measure gives an indication of the total blink length that occurred during the period when the participant rated the video clips as low risk or high risk normalised for the length of time they considered the clip segments to be low or high. (One participant had no rating data and another never rated any segment of the clips as being 'High', leaving 8 participants for this analysis). A repeated measures t-test comparing blink density during the High and Low risk sections of the clips was not significant, but the effect size (*d*) indicated a strong relationship (and that the failure to meet the critical t value resulted from the low sample size); *t*(7) = 1.7, *p* = .13, *d* = 0.6.

4.2.4. Analysis of pupil size and saccades. For each of the 20 video clips, the mean pupil size (n = 10 participants) and mean number of saccades during each video clip were calculated and, similar to the eye blink analysis, these measures were compared to the average risk rating for each clip (over the total length). The results of this analysis are shown in Figure 18. As can be seen in the figure, participants' pupil size and number of saccades were both positively correlated with ratings of risk; Pearson's $r(19) = .531$, $p = .016$ for pupil size and $r(19) = .737$, $p < .001$ for number of saccades.

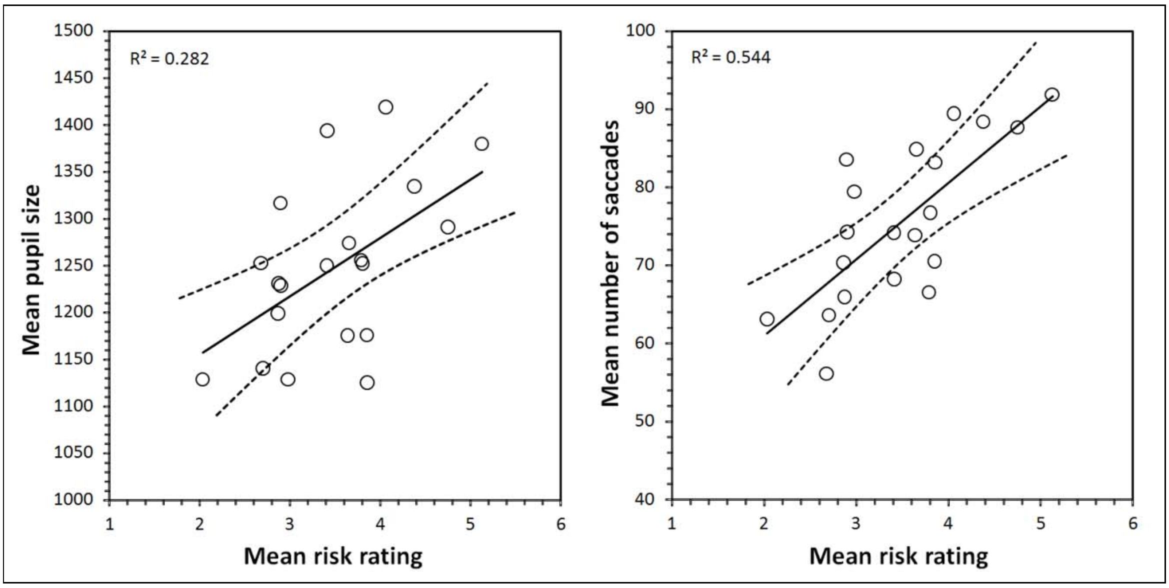

Figure 18. Mean pupil size (left) and number of saccades (right) for 20 video clips and the corresponding mean risk ratings. Dashed lines indicate 95% confidence intervals.

5. Naturalistic driving testing

This stage of testing was directed at measuring the subjective risk experienced while driving on a subset of the roads presented to participants in the driving simulator task. These risk ratings, and the participants' reasons for their risk ratings provided during the verification task, provided another point of comparison and cross-validation, as well as an explicit narrative to aid in interpretation of the data collected.

5.1 Method

5.1.1 Participants. A sample of 15 participants over 25 years of age, who held a full NZ driving licence, were recruited via flyers posted on noticeboards around the University, local community centres, shops and cafes. Electronic advertisements were also placed on the

University of Waikato's on-line learning platform (Moodle), Facebook and other social networking sites. Fourteen participants (8 males, 6 females, average age = 41.6 years, range 25-50 years, 10 of European descent and 4 'Other') completed all or part of the drive (the remaining participant was recruited to trial the procedure). Of the 14 participants who took part in the study, data was incomplete for three participants due to poor weather, other time commitments and equipment failure. The participants had been licensed drivers for 23.3 years on average (range 2-33 years). Six drivers reported that they had never been involved in a crash; eight had been involved in at least one crash. Three drivers had received an infringement notice in the past year.

5.1.2. Apparatus. A Suzuki SX4 vehicle, 2012 (see Figure 19, top left) was fitted with two video cameras (HD quality, equipped with a standard lens, f=55 mm); the first was attached via suction cups to the front windscreen, recording the road scene ahead (see Figure 19, middle panel); the second was attached to the small window on the passenger side of the vehicle to record the driver's behaviour (see Figure 19, bottom panel). A laptop computer (Figure 19, top right) controlled a program which generated a beep to prompt the participant to provide a verbal risk rating at predetermined GPS coordinates (points of interest) along the route. The computer also synchronised the video files from the two cameras and stored them.

Figure 19. The car, computer system, video cameras and screen shots (from the video cameras) from the naturalistic driving testing.

The participants drove a route containing a subset (13) of the test roads used in the driving simulator testing (see Figure 20). The route was selected as it was close to the research base at the University of Waikato, contained roads with a variety of features and hazards, and the drive could completed within a reasonable period of time (approx. 2 - 2.5 hr). The route began at the University leaving Hamilton along SH23 towards Whatawhata. Participants turned left onto SH39, Kakaramea Rd continued through Pirongia and then turned right on SH31, Kawhia Rd. Participants turned round at the junction of Kawhia Rd and Kaimango Rd and took the same route back to the university (approx. 180 km round trip). Participants were given a break and provided with a drink at the mid-point of the drive.

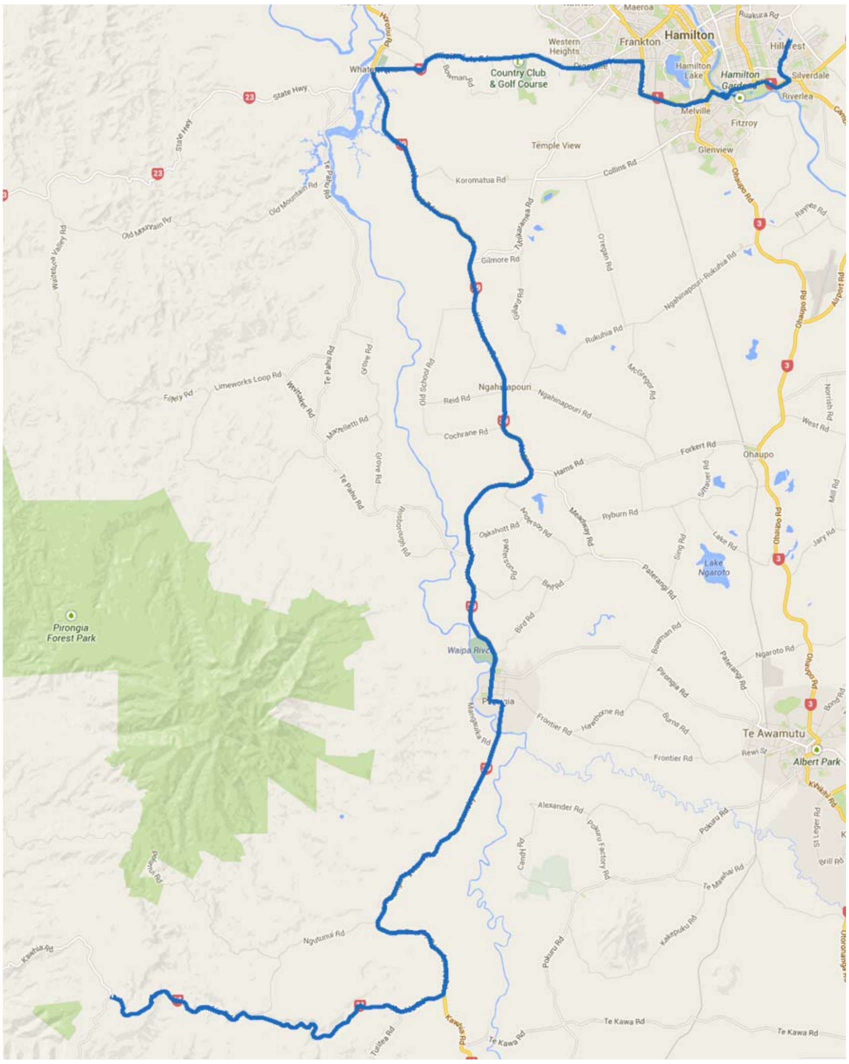

Figure 20. The route taken by participants during the naturalistic drive.

5.1.3. Procedure. After contacting the researchers, participants were sent an information sheet about the study. Participants were then invited to meet a researcher at a mutually convenient time to complete the drive. On arrival at the laboratory, the purpose of study was explained, the participants were shown a map of the route, any questions the participant had were answered, and the participant was then asked to sign the consent form. After this participants completed the demographic and driving history questionnaire and a copy of their driving licence was obtained. Participants were asked to comply with all normal roads rules and regulations and drive as they would in their own car. They were told about the scheduled rest-breaks and that they should let the research assistant know if they wanted to stop at any other time. They were also told that if they no longer felt comfortable driving the research assistant would drive them back to the University.

The participant was then taken on a short test drive (15-20 mins) in the instrumented car (along SH1B) to familiarise themselves with the vehicle and the tone that prompted them to provide a subjective risk rating (the risk rating scale this was the same as that used during the driving simulator and eye-tracking testing). After the test drive, the researcher provided instructions for main drive as follows:

> *The purpose of the study is to investigate the relationship between drivers' perceived levels of risk (that arise from the appearance of roads) and the objective levels of risk associated with those roads. You will drive around some Waikato roads, accompanied by me. I will provide you with directions and there is a map in the car. There are designated rest stops but if you need to stop at any other time just let me know. If you want to stop driving at any time just let me know and I will bring you back to the University.*
>
> *During the drive you will be prompted (by a beep) to provide verbal ratings of how risky you think certain parts of the road are (from 1 safe-10 very unsafe). If you felt completely at ease, if you were at rest or parked and could completely take your mind off driving - you would rate the risk as 1, but, if you felt extremely threatened, very unsafe, or in immediate danger of being involved in a serious accident or mishap you would rate the risk as 10. We want you to give us a rating between 1 and 10 every time you hear the beep.*

The verbal ratings of subjective risk provided by the participant were recorded (using pen and paper) by the research assistant in the vehicle in addition to being captured as part of the on-going video recording of each drive. At the end of the drive, participants received a $40 voucher to thank them for their time, and arrangements were made for them to come back to the laboratory to complete the verification testing (preferably within one week of the drive).

5.2. Results

The main aim of this part of the project was to determine the subjective risk experienced while driving on a subset of the roads presented to participants in the driving simulator task. The analysis examined the relationship between the mean risk ratings of the drivers who took part in the naturalistic study (14 participants) and the corresponding mean risk ratings provided by participants during the simulated driving testing (69 participants).

Figure 21 shows a scatterplot of the risk ratings for the 13 road segments, from the participants during the naturalistic drive with the corresponding risk ratings from the participants from the video-based driving simulator testing. As can be seen in the figure, there is a strong positive relationship between the risk ratings from the naturalistic drive and the video-based ratings; Pearson's $r(13) = .791$, $p = .001$.

The mean risk ratings were lower for the naturalistic drive (M = 3.44, SD = 0.97) compared to the ratings obtained for the same roads viewed in the driving simulator (M = 4.03, SD = 1.58). As shown in Figure 21, most data points are within, (or at) the 95% confidence intervals, except for the ratings of two test locations. For the first of these locations (narrow road, white guard rail on left and high bank on right), the naturalistic rating was lower than that predicted from the video-based risk rating. For the second location (straight road, narrow shoulders, poles), the naturalistic rating was slightly higher than predicted. The high correlation between the video-based ratings and those obtained from the naturalistic drive support the validity of the subjective risk data collected in the driving simulator sessions and suggest that these ratings correspond well to peoples' perception of risk on the road.

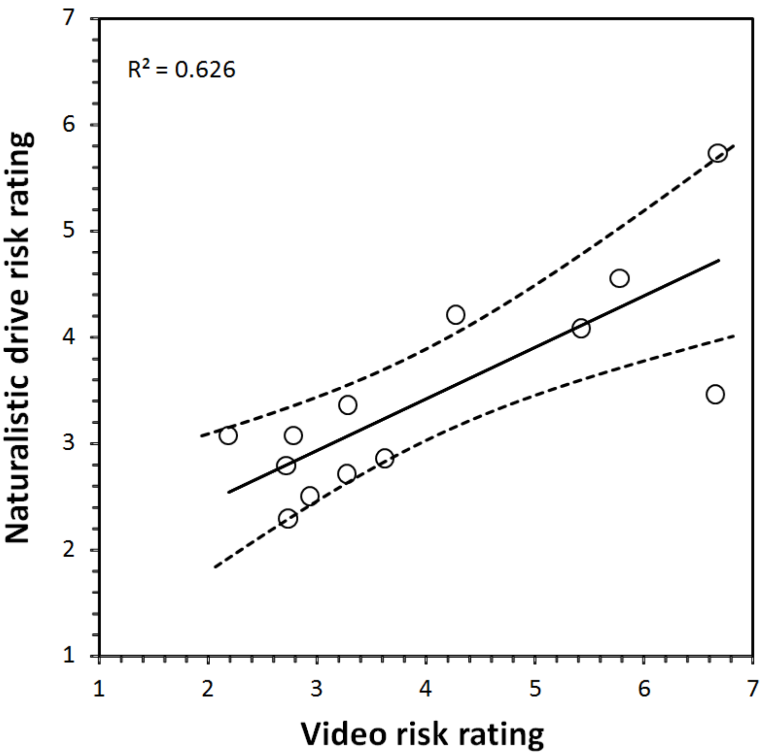

Figure 21. Mean risk ratings of the 13 test roads from the naturalistic drive and the video-based driving simulator ratings (1 = ‘safe’, 10 - ‘very unsafe’). Dashed lines show 95% confidence intervals.

6. Verification testing

During this stage of testing, the participants who completed the naturalistic drive viewed a video of their drive and provided an explanation about the risk rating they gave. This aim of this phase of testing was to provide verification of the on-road risk ratings provided by these participants and to aid in interpretation of the data collected in the driving simulator and with the eye tracker.

6.1 Method

6.1.1 Participants. The 14 participants who participated in the naturalistic driving testing also completed the verification testing.

6.1.2. Apparatus. Participants were seated in a comfortable chair and viewed the videos (from a computer) on a flat-panel display screen (93cm x 52 cm, 1920 x 1080 pixels) from a distance of 2.3m (Figure 22). For each participant, the sections of the video including the locations of the GPS initiated tone and risk ratings were identified from the recording of the entire drive. For each of the 13 locations a 35 second video clip was extracted (25 seconds before the tone to 10 seconds after) using Adobe Premier Pro software and used as individual stimuli for the verification testing.

Figure 22. A participant viewing video footage of their drive during verification testing.

6.1.3. Procedure. On arrival at the laboratory, participants were seated in a chair opposite the display screen. The researcher explained that they would be shown two practice video clips, followed by clips from their drive at the locations where they were prompted to provide risk ratings. At the end of each clip participants were given the opportunity to change the risk ratings they gave during their drive (or leave them unchanged) and also comment on road features that contributed to their ratings. The instructions to the participants were as follows:

I am going to show you some video clips of the drive we did the other day. Midway through the clip you will hear the beep and the rating you gave us at the time. We would like you to tell us what kinds of things you noticed that made you give that rating. Afterwards, if you want to change the rating you initially gave, feel free to do so, and tell us why you have changed it. We will start with two practice clips from another road.

Now I will show you the clips from your drive. Remember just tell us what kinds of things you noticed that made you give that rating. Afterwards, if you want to change the rating you initially gave, feel free to do so, and tell us why you have changed it.

Participants were initially presented with a black screen with the words “Ready to begin” in the centre, pressing the space bar on the computer initiated an on-screen 5 second countdown to the first of two practice video clips (taken from footage collected by the researchers). At the end of the first video clip, participants were presented with a blank screen whilst they provided a verbal response. When they were ready to proceed, the researcher depressed the space bar to initiate the presentation of the second practice clip, after a 5 second on screen countdown. Once participants understood the task, they viewed the 13 video clips from their drive in the same manner as the practice clips, that is each clip was preceded by a 5 second on-screen countdown and ended with a blank screen. Participants were allowed as long as necessary to provide a verbal response to each video clip.

The verification test sessions were recorded (audio and video) for subsequent analysis. Each session took approximately one hour. At the end of the session participants were thanked for their time and provided with a $10 voucher.

6.1.4. Analysis. The audio/video recordings of each verification session were viewed and the reasons provided for each risk rating were typed verbatim into a spreadsheet. The comments were reviewed and post-hoc categories (curves, visibility, traffic, terrain, narrow road, signs, straight road, bridge, road markings, junction, banks, weather, no shoulder) were derived by two scorers. The comments were assigned to the appropriate category regardless of whether participants used the particular road feature to justify a high or low risk rating, as the focus was on identifying features than informed risk ratings generally, rather than focusing on high risk features only. Once all responses had been coded, a count was performed for each category.

6.2. Results

The mean risk ratings from each point of interest on the naturalistic drive (M = 3.44, SD = .97) were remarkably similar to those obtained from the same participants during the verification testing (M = 3.41, SD = .97). A paired samples t-test confirmed that there was no significant difference between the on-road and video-based ratings, $t(12) = .50$, $p = .63$, suggesting that the level of risk experienced was similar across both presentation modalities.

In regard to the features that participants reported as contributing to their risk ratings (Figure 23) the most commonly mentioned related to curves (e.g., *swerving corner),* visibility (e.g., *can’t see ahead*), traffic (e.g., *idiot on motorbike; oncoming traffic)* and terrain (e.g., *brow of hill*). The road width was also noted, as were speed advisory signs, particularly heading into curves (e.g., *35 km/h sign close to corner*). All of the participants commented that the presence of the single lane bridge influenced their risk ratings (it was only present in one clip). Road markings (e.g., *yellow lines; no overtaking*), junctions, banks, weather and the lack of a shoulder were mentioned less frequently. The presence of poles and ditches did not appear to influence the participants' risk ratings.

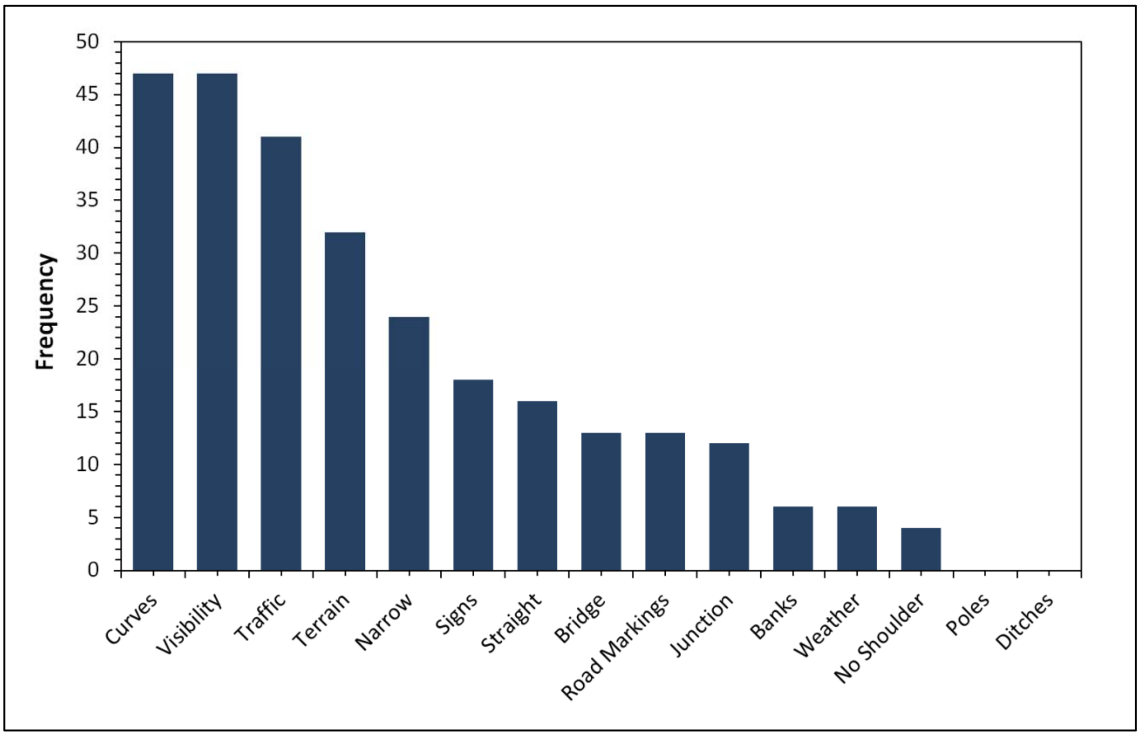

Figure 23. The frequency of reporting of specific road features as contributing to participants' risk ratings.

7. Discussion

Returning to the three research questions identified at the beginning of the project, the present study provides some useful answers to all three.

As regards the first question, "*What levels of subjective risk are experienced on hazardous New Zealand roads?*" the findings clearly indicate that drivers do perceive and consider risk while driving on New Zealand roads. The level of risk experienced is, in general terms, commensurate with the objective risk for the roads. In other words, roads with high objective risk are correctly perceived as high risk and roads with low objective risk are perceived as low risk by drivers. There are, however a few notable exceptions to this correspondence between perceived risk and objective risk that will be described in greater detail below.

For Question 2, "*What road features do drivers use to judge driving risk?"* the analysis of risk ratings from the driving simulator revealed that drivers use curves, hills, road width, and the presence or absence of a divided median to judge the risk of a road. These features alone predict nearly 80% of the variation in the participants' ratings and, notable by their absence, none of the other road features making up RPS scores contributed to drivers' perceptions of risk. Horizontal curves and narrow lanes were rated as the road conditions with the highest risk, and in many cases the perceived risk at these locations were higher than would be merited by the objective risk. Interestingly, the presence of wire rope barriers and bridge rails tended to increase participants' ratings of perceived risk, particularly for the still photos in the case of bridge rails. The data collected during the naturalistic drive supported these findings, with curves, terrain and road width being three of the most commonly reported features that influenced their risk ratings.

Finally, the results are unambiguous in answering Question 3 *"What hazardous road situations are under-recognised by New Zealand drivers (i.e., show the greatest dissociation between objective and subjective risk)?"*; intersections, and roadside hazards such as ditches and poles were significantly under-rated as risks, and the eye-tracking data indicated that participants do not even look at these roadside objects. The comments from the verification testing bolstered this finding in that none of the participants mentioned these roadside hazards as contributing to their ratings. Intersections were mentioned in some cases, and were rated as moderately risky, but not to the levels indicated by the objective risk associated with these locations.

The use of converging and complementary methods in the present research provides a high level of confidence in the above answers. The wide range in risk ratings across different road conditions and locations indicated that the roads selected for testing represented a good range of road features to be assessed by the participants. The consistency checks built into the driving simulator testing indicated that the participants' ratings were extremely reliable and reflected the perceptions of risk associated with the properties of the roads rather than random variation. Further, the remarkable consistency of results from the different testing methods (video, photo, driving simulator, eye-tracking, naturalistic driving, and verification tests) provides a level of assurance that the comparison of drivers' risk perception to objective risk is accurate and definitive. In fact, the present study represents the first time these different approaches to testing have been directly compared using the same stimuli and rating scale.

As such, the findings will make a valuable contribution to the published literature on the perception of risk.

The use of multiple testing methods not only allowed us to assess the reliability of the findings, they also provided a more complete picture of risk perception than has been previously possible. The pattern of risk ratings from the naturalistic driving testing supported the validity of the video and photo data collected in the driving simulator sessions, and the finding of consistently lower absolute ratings while driving suggests several new research questions to investigate. For example, did the lower levels of risk reported correspond to a lower experience of risk due to the competing attentional demands driving a car? Or did a greater sense of control lead the naturalistic drivers to feel more confidence and less risk?

Similarly, the verification testing supported the accuracy of risk ratings obtained from the driving simulator and on-road testing, but it also provided explicit insights into the reasons behind the participants' risk ratings. The participant reports from the verification testing echo the findings from the multiple regression analysis (from the driving simulator data), with curves (horizontal alignment), terrain and lane width playing a significant role in estimations of subjective risk. These reports, taken together with the results from the driving simulator and eye tracking testing, provide convincing evidence that narrow shoulders, poles and ditches tend to be under-rated and are not used explicitly to inform drivers' subjective ratings of risk.

The study also identified some potentially useful new measures of risk perception. For example, the finding that pupil size and the number of saccades increased as the participants' perceived risk increased, suggests that these measures might be used as a useful addition to other measures of perceived risk. Further investigation of these measures is warranted, however, using a somewhat different stimulus set to verify that the saccades and changes in pupil size were not simply due to the presence of curves, which the participants tended to rate as riskier than other road features. Saccades tend to be more frequent on sections of road with curves as the drivers tend to scan the bends and edges of the road more often. In the present study, this may have produced more saccades for high risk locations compared to the straight (low risk) sections of road where eye fixations tend to be concentrated in a smaller area close to the centre line and further in the distance. Recent evidence has also suggested that pupil size may be related to motor preparation for saccades (Jainta, Vernet, Yand & Kapoula, 2011) and this may explain why pupil size was also positively related to the risk rating. Given that more saccades occurred for the high risk (curved) road sections, it is possible that the participants' pupil size dilated prior to the saccades and thus generated a higher average pupil size for the high risk compared to the low risk roads. Some additional testing of these measures (pupil size and saccades) need to be carried out to see how robust they are and where they are most useful.

Additional research with the naturalistic driving and verification testing procedures would also be of considerable benefit. Due to budgetary and time constraints, data were collected from a relatively small sample of drivers and for only a subset of the roads used in the driving simulator testing. The effect of this means that the sample of drivers in this portion of the study may not fully represent the experiences and perceptions of the wider population of New Zealand drivers. Further, the frequency counts of the participants' reasons for their risk ratings do not take into account the unequal exposure to the various road features. For example, the limited number of times some road features were accompanied by a prompt for a rating during the naturalistic drive (e.g., bridges) limited the number of times it could be reported as a reason for participants' risk ratings. In contrast, changes in horizontal alignment and terrain were present during many more occasions and thus could inflate the frequency with which these road features were listed as reasons for the participants' risk ratings.

The driving risk associated with intersections could also usefully be explored in further research. The perceived risk of intersections in the present study was substantially different than their objective risk. Participants did, however, mention intersections (particularly those with other vehicles present) from time to time during the verification testing. Unlike roadside ditches and poles, which were not mentioned by the participants, the risks associated with intersections were under-rated rather than simply not considered. Exploring how to make the risks associated with intersections more apparent to drivers, or what aspects of intersections contribute to drivers' risk perceptions would be a valuable contribution to road safety. It is possible that additional work in this area might inform the KiwiRAP rating system by taking into account how perceived risk contributes to objective risk. As noted by earlier researchers, roads with low perceived risk may actually contribute to higher levels of objective risk because drivers are not taking due care at these locations. For example, it should be possible to apply the regression equation used to describe drivers' risk ratings in the present study to other sections of state highway and predict other locations where drivers' perceptions of risk might be dangerously low.

In conclusion, the present study was highly successful in answering the three research questions posed at the outset. The research approach allowed a very high level of confidence in identifying the levels of risk perceived by New Zealand drivers and locations in which the risk is under-rated. New testing methods, and the combination of methods, also revealed new insights about how driving risk is experienced and what road features contribute to it. Finally, as with any successful study, the present research suggests several important new questions and areas of future investigation.

8. Acknowledgements

The authors would like to acknowledge the patience and hard work of the participants, research assistants, and technical support staff who worked with us on this project. Without their dedication this work would simply not have been possible.

8. References

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*, 45-60.

Assum, T., Bjornskau, T., Fosser, S., & Sagberg, F. (1999). Risk compensation – the case of road lighting. *Accident Analysis and Prevention*, *31*, 545-553.

Benda, H. V., & Hoyos, C. G. (1983). Estimating hazards in traffic situations. *Accident Analysis and Prevention*, *15*, 1-9.

Brookhuis, K. A., de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, *42*, 893–903.

Burdett, B. R. D., & Nicholson, A. J. (2010). *Speed Management on Rural Roads*. In Proceedings of IPENZ Transportation Group Conference. Christchurch: Institution of Professional Engineers New Zealand.

Charlton, S. G. (2006). Conspicuity, memorability, comprehension, and priming in road hazard warning signs. *Accident Analysis and Prevention*, *38*, 496-506.

Charlton, S. G. (2011). *Improving Driver Awareness of Road Risk and Driver Behaviour Using KiwiRAP Ratings.* TERNZ Research Report. Auckland: Transport Engineering Research NZ.

Charlton, S. G. and Baas, P. H. (2006). *Speed Change Management for New Zealand Roads*. Land Transport New Zealand Research Report No 300. Wellington, NZ: Land Transport New Zealand.

Charlton, S. G. and Starkey, N. J. (2011). Driving without awareness: The effects of practice and automaticity on attention and driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*, 456-471.

Cooper, D. R. C., Jordan, P. G., Young, J. C. 1980. The Effect on Traffic Speeds of Resurfacing a Road. TRL Report 571. Crowthorne, Berkshire, UK: Transport and Road Research Laboratory.

Elvik, R. (2004). To what extent can theory account for the findings of road safety evaluation studies? *Accident Analysis and Prevention*, *36*, 841-849.

Evans, L. (1985). Human behaviour feedback and traffic safety. *Human Factors*, *27*, 555-576.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, *37*, 461-472.

Gibson, J. J, Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. *The American Journal of Psychology*, *11*, 453-471.

Groeger, J. A., & Chapman, P. R. (1996). Judgment of traffic scenes: The role of danger and difficulty. *Applied Cognitive Psychology*, *10*, 349–364.

Harless, D. W., & Hoffer, G. E. (2003). Testing for offsetting behaviour and adverse recruitment among drivers of airbag-equipped vehicles. *The Journal of Risk and Insurance*, *70*, 629-650.

Howarth, C. I. (1988). The relationship between objective risk, subjective risk and behaviour. *Ergonomics*, *31*, 527-535.

Jainta, S., Vernet, M., Yang, Q. & Kapoula, Z. (2011). The pupil reflects motor preparation for saccades- even before the eye starts to move. *Frontiers in Human Neuroscience*, *5*, 1-10.

Joshi, M. S., Senior, V., & Smith, G. P. (2001). A diary study of the risk perceptions of road users. *Health, Risk & Society*, *3*, 261-279.

Kanellaidis, G., & Dimitropoulos, I. (1994). Subjective and objective evaluation of risk on roadway curves. *Traffic Engineering and Control*, *55,* 451–454.

Kanellaidis, G., Zervas, A., & Karagioules, V. (2000). Drivers' risk perception of road design elements. *Transportation Human Factors*, *2*, 39-48.

Leden, L., Hämäläinen. O., & Manninen, E. (1998). The effect of resurfacing on friction, speeds and safety on main roads in Finland. *Accident Analysis and Prevention*, *30*, 75-85.

Lewis Evans, B. and Charlton, S. G. (2006). Explicit and implicit processes in behavioural adaptation to road width. *Accident Analysis and Prevention*, *38*, 610-617.

Liston, B.D., Krukowski, A.E., & Stone, L.S. (2013). Saccade detection during smooth tracking. *Displays*, *34*, 171-176.

Lund, A. K., & O'Neill, B. (1986). Perceived risks and driving behaviour. *Accident Analysis and Prevention*, *18*, 367–370.

McKenna, F. P. (1987). Behavioural compensation and safety. *Journal of Occupational Accidents*, *9*, 107–121.

Näätänen, R., and Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. *Accident Analysis & Prevention*, *6*, 243-261.

Pelz, D. C. and Krupat, E. (1974). Caution profile and driving record of undergraduate males. *Accident Analysis and Prevention*, *6*, 45–58.

Rothengatter, T. (2002). Drivers' illusions - no more risk. *Transportation Research, Part F: Traffic Psychology and Behaviour*, *5*, 249-259.

Sagberg, F., Fosser, S., and Saetermo, I., F. (1997). Investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers. *Accident Analysis and Prevention*, *29*, 293-302.

Taylor, D. H. (1964). Drivers' galvanic skin response and the risk of accident. Ergonomics, 7, 439-451.

Thomas, J. A. (2011). *"Hands On" Assessment for the Effect of Better Road Delineation on Driving*. Paper presented at the The New Zealand Roadmarkers Federation Conference 2011. Rotorua: Zealand Roadmarkers Federation.

Thomas, J. A., & Walton, D. (2007). Measuring perceived risk: Self-reported and actual hand positions of SUV and car drivers. *Transportation Research Part F. Traffic Psychology and Behaviour*, *10*, 201–207.

Walton, D., & Thomas, J. A. (2005). Naturalistic observations of driver hand positions. *Transportation Research Part F. Traffic Psychology and Behaviour*, *8*, 229–238.

Watts, G. R., & Quimby, A. R. (1980). *Aspects of Road Layout That Affect Drivers' Perception and Risk Taking*. TRRL Report 920. Crowthorne, Berkshire, England: Transport and Road Research Laboratory.

Wilde, G. J. S. (1982). The theory of RHT: Implications for safety and health. *Risk Analysis*, *2*, 209-226.

Wilde, G. J. S. (1988). Risk homeostasis theory and traffic accidents: Propositions, deductions and discussion of dissension in recent reactions. *Ergonomics*, *3*, 441-468.

Wilde, G. J. S. (2002). Does risk homeostasis theory have implications for road safety? *British Medical Journal*, *324*, 1149–1152.

Wong, Y. D., & Nicholson, A. (1992). Driver behaviour at horizontal curves: Risk compensation and the margin of safety. *Accident Analysis and Prevention*, *24*, 425-436.