Driver Interaction with a Traffic Light Assistant App: A Naturalistic Investigation

Kyle WILSON[1], Karl BRIDGES[2], Paul WARD[1], Simon PARKINSON[1], Tyron LOUW[3], Ryan COONEY[4]

[1]*University of Huddersfield, United Kingdom*
[2]*HFEX Ltd., New Zealand*
[3]*University of Leeds, United Kingdom*
[4]*Christchurch Transport Operations Centre, New Zealand*

Abstract.
A smartphone-based traffic light assistant application, EnLighten, was investigated for its ability to improve subjective driving experience and safely reduce the time it took drivers to ‘move off’ at signalled intersections. Five drivers participated in four trials over a period of three weeks. Testing took place on public roads in unaltered normal traffic conditions. It was found that EnLighten can reduce move-off times, however it presented safety risks. Advantages and disadvantages of naturalistic in-vehicle testing are also discussed.

Keywords. Driving; Naturalistic; Traffic congestion; Cognitive workload

1. Introduction

Traffic congestion is a cause of major frustration for road users and has negative environmental impacts through increased CO_2 emissions (Barth & Boriboonsomsin, 2008; Hennessy & Wiesenthal, 1999). One factor that influences junction capacity and network congestion levels is the time that drivers take to move away (hereafter referred to as ‘move-off time’) at signalled intersections after signals change from stop (i.e., red light) to go (i.e., green light) (Older, 1963). In-vehicle distractions have increased with advances in vehicle technology as well as the proliferation of personal electronics such as smartphones (Foss & Goodwin, 2014), which are frequently used in non-driving related tasks. Distractions such as these may often be responsible for drivers being temporarily unaware of signal changes and thus departing more slowly from intersections (Strayer & Johnston, 2001).

A transport operations centre in a major New Zealand city wanted to investigate whether a smartphone-based traffic light assistant application (EnLighten, 2015) had the potential to alleviate network congestion. In the current investigation, EnLighten was tested to determine whether the time taken by drivers to move-off at signalled intersections could be safely reduced. Of specific interest to this study, this application (app) provides the user with information about traffic signal activity such as the time remaining before a signal changes (e.g., time-to-green) and prepare-to-go alerts. Findings from trials in Taiwan suggest that providing time-to-green information can reduce the number of crashes but without improving (i.e., reducing) move-off time (Chen, Zhou, & Hsu, 2009). It is possible that the mode of information presentation (i.e., using countdown timers external to the vehicle, installed at intersections) may have affected the result in this study, and a reduction in move-off time may occur when this information is provided in-vehicle via an alternative medium (i.e., a smartphone app). The purpose of this paper is twofold: First, to report preliminary findings on the effectiveness of a smartphone app providing drivers with traffic signal timing information, and second, to describe a method for highly naturalistic human factors testing in a light vehicle domain.

2. Methods

2.1 Participants

Five adults (3 females, 2 males) with a mean age of 33 years (SD = 9 years) participated. All had at least six years driving experience. Selection criteria included possession of a current driver's licence, current residency within the city where testing took place, current frequency of driving in that city (i.e., minimum of three days per week over the past year and over the upcoming trial period), preparedness to use their personal smartphone during driving, and having an Android phone (the platform on which the apps were run). Participants were recruited via word of mouth or social media and directed to an online questionnaire for screening purposes. Participants were compensated with a $50NZD fuel voucher.

2.2 Materials

2.2.1 EnLighten application. This 'app' provides two pieces of information about nearby traffic signals. First, the app presents a visual (second-by-second) countdown of the time remaining before the signal will change from Stop to Go (i.e., 'time-to-green'; see Fig 1). The visual countdown disappears 5 seconds before the green traffic light is illuminated. An auditory alert is subsequently presented shortly prior to the green, to signal to the driver to 'prepare-to-go'. Second, an auditory alert is provided as the driver's vehicle approaches a signalled intersection to provide information about the likelihood that the driver will have to stop. Low likelihood of having to stop is indicated by a higher pitch, 'cheery' tone, while a high likelihood of having to stop is indicated by a lower pitch, more sombre (i.e., less-cheery) tone. No tone is presented when the odds of having to stop were essentially equal (and the decision about stopping more uncertain). EnLighten utilises the device's GPS and mobile data in order to communicate with the traffic signals, and requires the city to have an integrated network of smart traffic signals.

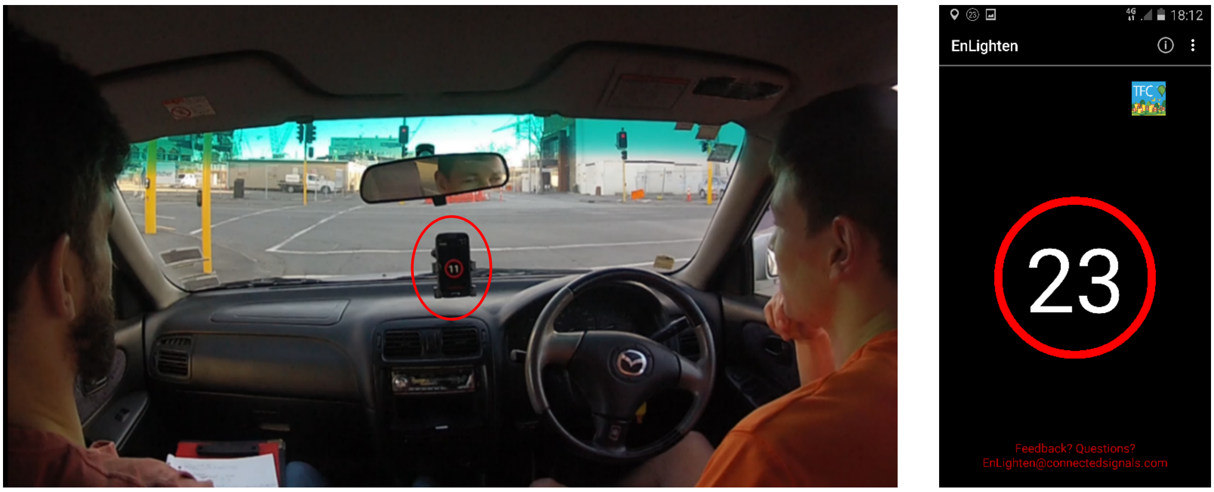

Figure 1. (Left) Photo depicting a driver using EnLighten, with a researcher in the passenger seat. Note that this image was taken during pilot testing, without all video recording equipment in place. (Right) A screenshot of the EnLighten app, showing a situation where 23 seconds remain before the traffic signal changes to go (time-to-green = 23 s).

2.2.2 Naturalistic trials. All participants completed four trials in their own personal vehicles. Approximately 75% of the route was pre-determined while the remaining 25% was determined by the driver during each trial. This was done to try to achieve a balance between control and realism. The pre-determined section was the same for all trials. The route entailed a drive-time of approximately 25-35 min over a distance of 16 km (10 miles), and included approximately 20 signalled intersections. Participants were not instructed to position or mount their cell phone in any way, they were simply told to position the phone in the way that they preferred. Four of the five drivers used cell phone mounts attached to the dashboard or windscreen, while one participant opted to place the phone on the centre console. The experimenter rode with the participant and passively engaged in conversation with the participant rather than attempt to stifle conversation since pilot work had indicated that this avoided adding unnecessary participant tension and stress, and better simulated natural driving conditions. At the end of trials the participants were instructed to safely stop. Each of the four trials was conducted approximately one week apart.

2.2.3 Measures. GoPro Hero4 high-definition video cameras were used to record both the internal (facing towards participant) and external (facing frontwards; driver point-of-view) environment. Video was recording at 60 frames per second with a viewing angle of 170 degrees. The video was used to capture key events (e.g., participant difficulty with the app, participant comments relating to the apps, etc.) and to measure the time taken to move-off at intersections following a change-to-green. A voice recording device was also used to record all interviews to supplement researcher notes made during trials.

The time it took drivers to move-off ('move-off time') was defined as the time between when a traffic signal changed to green and when the vehicle began to move from a halt. This was measured using a frame-by-frame analysis of the video and then converting frames to milliseconds, resulting in a margin of error of 16.67 ms (i.e., one frame length). Move-off time was calculated for all incidents during a trial where a participant stopped at a signalled intersection. To prevent any possible impact from biases, analyses were blinded so that the researcher was unaware of whether they were analysing a trial where the app was present or absent.

Several questionnaire measures were used. Driver workload was gauged using the Driving Activity Load Index (DALI; Pauzié, Manzan, & Dapzol, 2007). This is a driving-specific measure of workload containing five Likert-scale items scored from 0 ("Low") to 5 ("High"). Situation awareness was measured using the Low Event Task Subjective Situational Awareness measure (LETSSA; Rose, Bearman, & Dorrian, 2012), which is based on Endsley's (1995) model of situation awareness. This contained ten Likert-scale items scored from 0 ("strongly disagree") to 10 ("strongly agree"). Finally, a measure of app usability was taken using the Subjective Usability Scale (SUS; Brooke, 1996). This contained ten Likert-scale items scored from 0 ("strongly disagree") to 5 (strongly agree").

Semi-structured interviews were conducted immediately after each trial and were based on a set of pre-determined questions. At the beginning of the third and fourth trials, participants were also asked about their experiences with the app over the previous week (i.e., during personal travel when the researcher was not present).

2.3 Procedure

Participants completed an informed consent process prior to participation. Prior to each trial, the researcher set up the recording equipment in the driver's personal vehicle and initiated recording.

2.3.1 Baseline trial (trial 1). At the beginning of the first trial, participants were shown a map detailing the intended route. Once comfortable, the participant was instructed to begin the trial. This trial was conducted without using the app and used as a baseline to compare against subsequent test trials (i.e., 2 to 4). Once the participant arrived back at their start point, they completed the questionnaires (DALI and LETSSA) and were then interviewed.

2.3.2 App trials (trials 2–4). The second trial commenced by installing the application onto the participant's device (with their consent). The participant was then given a few moments to view the app before the task began. This differed to the procedure for trials 3 and 4, where participants were then informally interviewed about their experiences with the app over the past week, since their previous trial. The in-trial procedure for trials 2-4 was the same as for the baseline trial however. Following the trial, the same questionnaires given during the baseline were administered, with the addition of the SUS. The trials ended with a semi-structured interview.

3. Results

3.1 Move-off times

A comparison of the average move-off time in the baseline trial (app-absent) versus the average move-off time over the EnLighten trials (app-present) was initially performed with the data averaged over all five participants. However, given the unanticipated high rate of malfunctioning of the app, app-present incidences were further divided into incidences where the app worked (i.e., had given a prepare-to-go notification) and incidences where it did not work (see Fig. 2), which occurred 31% of the time. The number of move-off events on each trial naturally varied due to signal timings dictating whether a participant had to stop or not, and on average there were 4.5 move-off events captured on each trial.

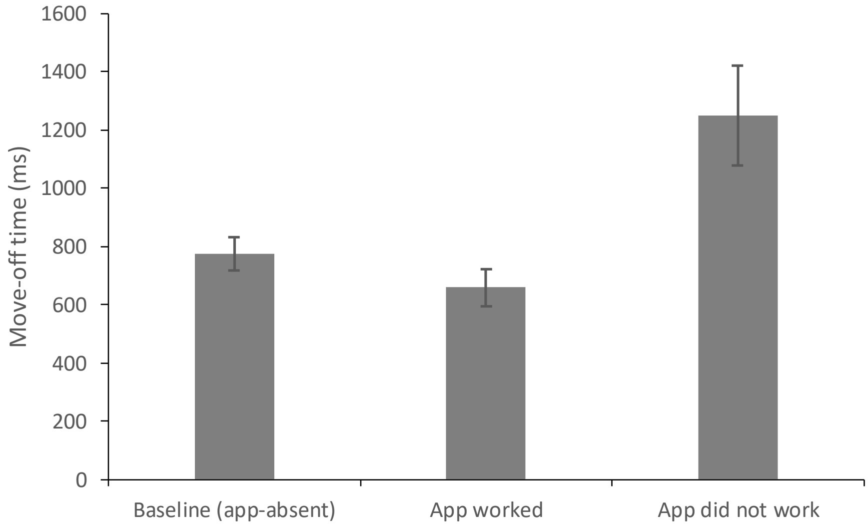

Figure 2. Length of time it took drivers to move-off after change-to-green. Error bars are standard error of the mean.

When the app functioned properly, participants were faster to move-off (M = 660 ms, SD = 439) compared to baseline trials where the app was absent (M = 775 ms, SD = 216). However, when the app did not function properly, participants were much slower to move-off (M = 1249 ms, SD = 784). When app-present trials were considered as a whole, participants actually took longer to move off (M = 842 ms, SD = 625) compared to baseline trials. Despite the small sample size we conducted null hypothesis significance testing. A paired samples *t*-test was used to compare move-off times for trials where EnLighten functioned correctly and for the baseline trials. There was no significant difference found, $t(4) = 0.81$, $p = .479$, $d = .40$. However, with only five participants this test is severely underpowered and with a larger sample size it is likely that significant differences would be found.

On 17% of incidences where participants were given a prepare-to-go alert, they began moving before the signal had changed green. These incidences were approximately evenly distributed across trials two to four. It should be noted that when all of the incidences where drivers moved-off early were removed from the analysis (these had been scored as '0 ms'), there was no longer an apparent time difference between the move-off times for baseline versus incidences where EnLighten worked.

3.2 DALI scores. Nonparametric tests were used to compare scores over each trial. Findings should still be treated with caution however. Table 1 displays the DALI scores over each of the trials. A Friedman test indicated that visual demand was rated differently over the trials, $\chi^2(3) = 11.40$, $p = .010$. It appears that the app reduced visual demand, particularly following one week of use. No significant differences were detected for the other subscales or for the DALI average, although a visual inspection of the global attention demand subscale and the DALI average suggests the app may have contributed to lower ratings for both of these.

*Table 1. Means and standard deviations for DALI scores over each of the trials. * p < .05.*

	App absent	App present		
	Trial 1 (Baseline)	Trial 2	Trial 3	Trial 4
Auditory demand	2.0 (1.4)	2.2 (1.8)	1.2 (0.8)	1.0 (1.0)
Visual demand*	3.2 (2.5)	2.8 (2.2)	1.4 (1.5)	1.6 (1.5)
Global attention demand	2.4 (2.4)	2.2 (1.6)	1.6 (1.5)	1.6 (1.5)
Interference (from app)	N/A	2.0 (1.3)	1.6 (1.5)	1.2 (0.8)
Stress	1.6 (1.5)	2.0 (1.6)	1.2 (0.4)	1.6 (1.5)
DALI average	2.3 (1.8)	2.2 (1.6)	1.4 (1.0)	1.4 (1.2)

3.3 LETSSA. A nonparametric test was used to compare ratings over trials. While a visual inspection of the averages for each trial suggests that situation awareness was improved by the introduction of the app, with the baseline mean average of 78.4 increasing to 87.1 (for trials 1-3 combined), no differences were detectable with a Friedman test, $\chi^2(3) = 6.06$, $p = .11$.

3.4 SUS. Usability scores varied considerably between participants as well as between trials. This appeared to be, unsurprisingly, closely tied to the rate of malfunctioning of the app. On several trials throughout the experiment, there were issues with participants' devices (or the app; this was not totally clear) losing GPS connectivity which was a frequent source of frustration for participants. Looking at the scale items individually revealed that the biggest detractor for SUS scores was the item relating to inconsistency: "I thought there was too much inconsistency in this app". The item which yielded the highest score was the perceived usefulness of the information. Further examination of the SUS subscales suggested that participants were able to learn how to use the app reasonably quickly and found it easy to use. Overall, the average usability score was 71.4 (SD = 12.6), which is acceptable when compared to existing product standards (Bangor, Kortum, & Miller, 2008).

3.5 Interviews. In the interests of brevity, only a few key findings will be described here. The most positive findings were that participants enjoyed using EnLighten, found the information to be useful, and believed that the time-to-green countdown reduced their stress levels. While perceptions of the app were typically highly positive on the first trials, participants became frustrated with the app as they encountered unreliable alerts and connectivity problems over their two-week usage periods. Participants felt that the biggest risks associated with using the app arose when it was malfunctioning, as this led them to focus on the app and 'fiddle' with it while driving. When users thought the app was not working they reported doing several things (all while driving) such as checking the settings, restarting it, and turning the volume up on their device. Observations by the researchers during the trials revealed the same behaviours.

One participant reported being caught off guard several times while waiting to turn at intersections, where EnLighten indicated a green signal but the participant could not have turned due to oncoming traffic. EnLighten often gave inadequate signal information on occasions where a driver intended to turn rather than proceed straight through signals. This seems to be because there is no easy way for the app to detect whether the driver intends to turn or to go straight. When turning, there is sometimes a requirement to yield to oncoming traffic, or pedestrians, which can take priority over the green signal.

Participants found the auditory alerts regarding whether the driver was going to have to stop at upcoming signals useful, although they felt that they did not occur frequently enough. At the time however, participants were not aware that the app only gave alerts when the predictions could be made with a reasonable degree of certainty. Users' perceptions of this feature may become more positive if this information is provided to them.

Other comments on Enlighten included: The requirement to use cellular network data (which typically has associated costs) was a drawback, and the app appeared to drain some users' phone batteries quicker than expected.

4. Discussion and Conclusion

Providing drivers with time-to-green information and prepare-to-go alerts appeared to reduce drivers' move-off times. Frequent reliability issues were encountered with EnLighten however. When it malfunctioned, it had the opposite effect to what was desired, leading to longer delays in move-off times. Throughout the current trials, the rate of malfunctioning was high enough that trials with the app resulted in slower move-off times overall. It could be that drivers were relying on the app to produce a prepare-to-go alert and therefore they devoted less attention to the actual traffic signals at the intersection. Alternatively, drivers may have become aware that the app had not given the prepare-to-go alert and as a result may have been distracted with it. Either way, it suggests less attention was paid to the external traffic signals when the app malfunctioned. This also demonstrates the negative influence of unreliable technology.

Another important finding and a potential concern was that during the trials with EnLighten, on 17% of the time drivers were stopped on a red signal they began to move-off early before the signal had changed to Green. Before testing, there was a concern that this might happen due to the timing of the time-to-green alert. As noted above, shortly before the signal changes to green the visual countdown disappears and an auditory 'prepare-to-go' alert sounds. This alert is intended to serve as a *prepare*-to-go cue, as opposed to a "go now" cue, however, participants reported occasionally mistaking the alert for a "go now" cue. These errors could be due to inexperience with the app, distraction, ambiguity of the alerts, or a combination of these factors although this is speculation. On these occasions, participants appeared to realise their mistake quickly and halted again. On other occasions, early move-offs appeared to be deliberate, with participants taking advantage of the prepare-to-go alert and getting a 'head start' so-to-speak. On the vast majority of these cases, however, the vehicle had *not entered* the intersection before the change-to-green (although this is based on subjective judgements by the researchers, regarding the precise location of the vehicle on the road), and therefore would not be classed as illegal. This was despite the vehicle having begun moving before the change-to-green. Note that we did observe one isolated incident where it was technically illegal and unsafe. Further investigation is required to determine how much of a safety risk this behaviour poses. Whether this behaviour would develop for other traffic light assistant systems, which also give time-to-green information, is not currently clear. It is also not clear whether this behaviour would increase or decrease with prolonged use, over and above the two-week period used here.

Subjective ratings of EnLighten were generally positive with participants reporting reductions in workload and stress when using it. The reduction in visual demand could be due to participants spending less time watching for traffic signal changes, given that the app informs participants of the time-to-green and provides a prepare-to-go cue shortly before the signal changes to green. Reports of situation awareness levels suggest that either these were not impacted, or marginally increased.

The naturalistic methods employed for this research had both advantages and disadvantages. Conducting testing on public roads with unaltered traffic flows exposed participants to conditions that were extremely close to those they would likely be experiencing day-to-day, as well as offering the same consequences. These are factors that can be difficult or impossible to replicate in simulators (de Winter, van Leeuwen, & Happee, 2012; Kappler, 1993) or even test tracks, and that likely have a considerable impact on drivers' interaction with in-vehicle information systems. The applied nature of this sort of testing is time-consuming however and when combined with budget constraints this can lead to smaller sample sizes being tested, as was the case here. Given our small sample size, the findings here should be considered preliminary.

One limitation concerned the method we used to gauge the time it took drivers to move-off at intersections. This was judged by carefully counting the number of video frames that passed between a signal change and the vehicle first moving. There are undoubtedly more valid and reliable methods for this, however, this approach may be useful for naturalistic research when budget and resources are constrained. It should also be noted that the average move-off times were based on a small number of incidences (on average, 4.5 times per trial).

Finally, we are unable to tell how users' interactions with this app—as well as other traffic light assistant systems—might change with prolonged use, given we only tested participants up until two weeks of usage. It is possible that drivers could develop an inappropriate level of trust in the automation (see Lee & See, 2004) and rely too much on the information given by EnLighten (overreliance; see Parasuraman, Sheridan, & Wickens, 2000). This could manifest as a driver relying solely on the app's alert as a cue for when traffic signals change, as opposed to the traffic signals themselves. In an event where the app incorrectly provides a prepare-to-go alert, when in fact, the signal is not due to change, overcompliance (see Parasuraman et al., 2000) could lead to a driver entering an intersection on a red light. In contrast, the opposite error (where the app fails to alert the driver of an upcoming signal change) might cause the driver to take longer to move-off (reliance), however it seems likely that this type of error would present less of a safety risk. Another situation where overcompliance could arise is when a driver intends to turn at an intersection and has a green signal but has to yield to oncoming traffic travelling straight through. Finally, the fact that move-off times were considerably greater when EnLighten malfunctioned suggests that overreliance on the app developed in the short space of time that the trial period was held over. Future research is necessary to examine the long-term effects on drivers' behaviour of using traffic light assistant apps such as EnLighten. Attempts should also be made to consolidate findings with countdown timers installed at intersections which may not be have as great of an effect on move-off time (see Chen et al., 2009, for a discussion). It would also be useful to compare how driver interaction differs between timers based on smartphones (as was investigated here) and timers based on in-vehicle information systems. BMW's iDrive infotainment system with EnLighten integrated is one example of this (BMW Group, 2015), and Audi has recently launched a traffic light assistant system in their 2017 Audi Connect infotainment system (Audi Media Center, 2016).

In summary, EnLighten has the potential to effectively shorten the time users take to move-off at signalled intersections and, therefore, may have some use in managing traffic congestion levels. Furthermore, the app can reduce drivers' stress levels and improve their subjective driving experience overall. However, all of these effects are highly contingent upon the system functioning as intended, and the current investigation shows that if it does not, the opposite effects occur. The current research also highlights some benefits and drawbacks of using an ecologically valid approach to study driver behaviour.

References

Audi Media Center (2016). Audi networks with traffic lights in the USA [Press release]. Retrieved December 19, 2016, from https://www.audi-mediacenter.com/en/press-releases/audi-networks-with-traffic-lights-in-the-usa-7147

Barth, M., & Boriboonsomsin, K. (2008). Real-world carbon dioxide impacts of traffic congestion. Transportation Research Record: Journal of the Transportation Research Board (2058), 163-171.

Bangor, A., Kortum, P.T., & Miller, J.T. (2008). An empirical evaluation of the system usability scale. International Journal of Human–Computer Interaction, 24, 574-594.

BMW Group (2015). BMW Group Announces the First Integration of EnLighten App [Press release]. Retrieved December 19, 2016, from https://www.press.bmwgroup.com/usa/article/detail/T0227922EN_US/bmw-group-announces-the-first-integration-of-enlighten-app?language=en_US

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability Evaluation in Industry, 189, 4-7.

Chen, H., Zhou, H., & Hsu, P. (2009). What Do We Know About Signal Countdown Timers? ITE Journal on the Web, July, 72-76.

de Winter, J., Van Leuween, P., & Happee, P. (2012). Advantages and disadvantages of driving simulators: A discussion. In Proceedings of Measuring Behavior, 47-50.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37, 32-64.

EnLighten (2015). Connected Signals, Inc. (Version 4.6.0) [Mobile application software]. Retrieved December 19, 2016, from https://apkpure.com/enlighten-by-connected-signals/com.otsys.notifier

Foss, R.D., & Goodwin, A.H. (2014). Distracted driver behaviors and distracting conditions among adolescent drivers: Findings from a naturalistic driving study. Journal of Adolescent Health, 54, 50-60.

Hennessy, D.A., & Wiesenthal, D.L. (1999). Traffic congestion, driver stress, and driver aggression. Aggressive Behavior, 25, 409-423.

Käppler, W.D. (1993). Views on the role of simulation in driver training. In Proceedings of the 12th European Annual Conference on Human Decision Making and Manual Control, Kassel, Germany, 5.12-5.17.

Lee, J.D., & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 46(1), 50-80.

Older, S. J. (1963) Omission of the red/amber period at traffic signals. Traffic Engineering and Control, 11, 414-417.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30(3), 286-297.

Pauzié, A., Manzan, J., & Dapzol, N. (2007). Driver's Behaviour and Workload Assessment for New In-vehicle Technologies Design. In Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Stevenson: Washington.

Rose, J.A., Bearman, C.R., & Dorrian, J. (2012). Constructing and evaluating the low-event task subjective situation awareness (LETSSA) measure. In Proceedings of the 10th International Symposium of the Australian Aviation Psychology Association, Australia.

Strayer, D.L.,& Johnston, W.A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. Psychological Science, 12, 462-466.