

Victoria's new Drive Test: Development of the assessment framework

Catchpole JE¹, Cavallo A², Christie R³, Harrison WA⁴,
Johnston D⁵, Macdonald WA⁶, Oh S² & Triggs TJ⁷

¹ARRB Group; ²VicRoads; ³RCSC Services; ⁴Eastern Professional Services;
⁵DCS Consultants; ⁶La Trobe University; ⁷Monash University
email: john.catchpole@arrb.com.au

Abstract

As part of the introduction of a new Graduated Licensing System, VicRoads has developed and implemented a new on-road licence test that is calibrated to the driving aptitude of today's licence applicants, most of whom are required to have at least 120 hours of supervised driving experience. This paper describes the evolution of the assessment framework for the new Drive Test through a series of trials involving driving tasks and assessment items taken from Victoria's previous driving test; tasks and items from tests used in other jurisdictions; and tasks and items newly devised by the Drive Test development team. The final assessment framework includes credit for correct demonstration of safe driving skills and penalties for driving actions that are illegal and/or create immediate danger to any road user. Although a number of alternative scoring protocols were trialled, the final framework scores all assessment items as a simple Yes or No (plus the possibility of Not Assessable for some items). The assessment criteria for the various items and error categories seek to promote reliability of assessment through clear specification of objectively verifiable behavioural requirements. VicRoads is monitoring the performance of the new Drive Test and the LTOs who administer it. Further refinements will be made if necessary to ensure consistency of assessment across all licence testing centres in Victoria.

Keywords

Licence Test, Driving Test, Test Development

Background

Under Victoria's new Graduated Licensing System (GLS), introduced progressively over the last two years, most driver licence applicants are required to have completed at least 120 hours of supervised driving experience while holding a learner permit before taking the on-road driving test to obtain a probationary licence.

Victoria's former driving test, known as POLA (Programmed Licence Observation Assessment), was developed and introduced in the early 1990s, when the typical level of driving experience of learners applying for a licence was far lower than is expected of today's applicants. As part of the new GLS, VicRoads needed to develop a new on-road driving test better suited to the skill levels of today's more experienced applicants. The aim was not only to develop a test that is better calibrated against the aptitude of present-day applicants, but also, much more ambitiously, to develop a test that can discriminate between applicants who have accumulated the required 120 hours of supervised driving experience and those who have not.

The development process was evidence-based, with a series of major and minor trials being used to assess the suitability of various driving tasks and assessment protocols to meet the aims of the new test. The test components subjected to trialling included components of Victoria's POLA test; components borrowed from driving tests used in other jurisdictions; and tasks and assessment methods newly constructed by the test development team to examine critical driving skills.

This paper describes the process of developing and defining the assessment items used to score the applicant's performance in the new Drive Test. Other papers presented at this conference provide an overview of the new Drive Test and its role in Victoria's new GLS [1]; explain the conceptual basis of the

test [2]; describe the selection of the best performing items on the basis of data collected during the trials [3]; and provide an account of the implementation of the new test across Victoria [4].

Building blocks of the Drive Test

Most of the 'performance checks' (i.e. assessment items) in Victoria's former driving test were location performance checks, which assessed driving performance at pre-defined locations on the test route. Each location performance check required the applicant to demonstrate a range of correct behaviours in order to score a point. For example, performance check *J* (judgement) required the applicant to demonstrate appropriate speed control, check for approaching vehicles and pedestrians and select a safe gap; performance check *MSP* (mirrors, signalling, position stop) required the applicant to check the appropriate mirrors, signal in the appropriate direction for the appropriate period, stop at the appropriate position and demonstrate safe gap selection when moving off after stopping. If all of the required behaviours were correctly demonstrated, the applicant scored a point, but if any one of the requirements was not met, the applicant scored zero for the entire performance check.

For the new Drive Test, the development team decided that each assessment item should focus on the correct performance of a small set of closely-related skills, rather than the broad range of disparate skills relevant to a driving task. For example, *Observation* skills (visual search, mirror checks and head checks) are assessed as one item; *Signalling* is assessed as a separate item; and *Gap Selection* is another separate item. During test development, the use of separate assessment items for each skill group allowed the development team to assess which (if any) skill group was most closely related to driving safety and to prior driving experience and hence most worthy of inclusion in the final test.

Narrowing the range of skills assessed by a single item makes it possible to assess more than one item during the performance of a single driving task. During licence testing, this approach allows applicants to receive partial credit for partial compliance with the criteria for the task. For example, if an applicant making a left or right turn at an intersection fails to ensure the turn indicator is cancelled as soon as the turn is completed, he or she will lose a point for signalling but is still able to score a point for correct observation and another for safe gap selection. This not only supports finer discrimination of driving aptitude but also allows more precise feedback to be given at the end of the test regarding skills that still need improvement.

Thus a test drive comprises a number of assessable driving tasks. When each assessable task is undertaken, the Licence Testing Officer (LTO) assesses the applicant's performance of the task using the assessment items applicable to that task. A single assessment item, such as *Signalling*, may be applicable to several driving tasks, such as left turns, right turns and lane changes. A test drive also includes linking manoeuvres that join up the assessable tasks to form a continuous test route. Although no task assessment items apply specifically to the linking manoeuvres, it will be seen later that other forms of assessment apply continuously throughout the test drive, both during assessable tasks and linking manoeuvres.

Evolution of assessment items during the trials

Phase 1

During Phase 1 of trialling, assessment items from Part 1 of New Zealand's Full Licence Test (FLT), from California's Driver Performance Evaluation (DPE) and from Victoria's then-current POLA test were trialled for possible inclusion in the new Drive Test.

Part 1 of the New Zealand FLT includes eight different assessment items (*Mirror Use, Search, Signal Use, Head Check, Braking, Speed Control, Position* and *Gap Selection*). In the FLT, all items are scored on a Yes/No scale, with Yes being recorded if the applicant meets the requirements for the item and No if the applicant does not meet the requirements.

All eight FLT assessment items were trialled during Phase 1. Three of these items (*Mirror Use, Search* and *Head Check*) were scored on a Yes/No scale just as in the FLT. A further four items (*Signal Use, Braking, Speed Choice* and *Gap Selection*) were scored on a 0/1/2 scale that allowed three different levels

of performance to be recorded. For example, *Signal Use* was scored as 0 if the participating driver failed to signal when required; 1 if the driver signalled but started signalling too late or finished signalling too early; and 2 if the driver signalled at the appropriate time for the appropriate duration. The remaining FLT item, *Position*, was split into two items, *Stop Position* and *Lateral Position*, which were both scored on Yes/No scales.

In addition to the eight items taken from the FLT, two assessment items newly developed by the test development team were trialled during the FLT section of the Phase 1 trial routes. The new items were *Observation* and *Acceleration Smoothness*, both rated on an 11-point scale from 0 (very poor) to 10 (excellent). These two ratings were assessed once for each participating driver. They assessed performance over the entire FLT section of the trial drive, not just the assessable tasks. The rating items were included in the trial in order to investigate whether LTOs could provide useful information through subjective ratings that was not captured by Yes/No items with objective criteria.

In California's DPE, a test drive comprises six core sections and two optional sections (curve driving and freeway driving). All of the six core sections, together with the 49 assessment items that apply to those sections, were included in the DPE section of the Phase 1 trial routes. Most of the 49 assessment items have multiple behavioural criteria. Consideration was given to splitting compound assessment items into a number of simpler items, but this would have resulted in an impossibly large number of assessment items for the LTOs to learn and to assess during the 15-minute DPE stage of the trial drive. Allowing partial credit for satisfying some but not all of the criteria for an item was also considered but rejected due to the workload that would have been imposed on the LTOs. Thus the 49 DPE assessment items were trialled as defined by the developers of the Californian test.

POLA (Victoria's former driving test) included items assessed at specified locations along the test route (location performance checks) and items assessed over a specified section of the test route (stage performance checks). Apart from the low speed manoeuvre (such as a reverse park), only one location performance check was assessed at each location on the test route. However, some of these were combined performance checks that assessed several aspects of driving skill. Performance check *MSP* (mirrors, signalling, position stop), mentioned earlier, is an example of a combined performance check. In order to investigate the ability of each component performance check to discriminate between drivers with and without the required 120 hours of supervised experience, it was necessary to record separate assessments for each component performance check, rather than a single assessment for the combined performance check. Thus during the trial not just one but several POLA performance checks were assessed and recorded at each assessable location on the test route.

For each assessable task in all three sections of the trial route (FLT, DPE and POLA), the score sheet provided spaces for the LTO to record the result of each assessment item and spaces to record whether the applicant stalled the vehicle and whether any intervention by the accompanying driving instructor or LTO was required to complete the task safely.

Phase 2

During Phase 2 of trialling, assessment of *Observation* and *Smoothness* was applied to individual driving tasks. This permitted the collection and analysis of data on variations in these aspects of performance between different types of driving task. This information was not available in the Phase 1 trial, where these two items were assessed over an entire stage of the test route. *Observation* was assessed on a simple 0/1 (No/Yes) scale, while *Smoothness* was rated on a 5-point scale from 0 (satisfies none of the stated criteria) to 4 (satisfies all of the criteria).

The development team considered it possible that LTOs may be able to make useful judgements about driving performance based aspects of behaviour not addressed by the item assessment criteria developed by the team. To investigate this possibility, LTOs in the Phase 2 trial were asked to rate the *Quality of Driving* over the entire test route on a scale from 0 (very poor for a licence applicant) to 10 (very good for a licence applicant).

New assessment items were devised for Phase 2 of trialling to assess performance on several new driving tasks devised by the test development team. The new tasks were:

- the secondary task (described elsewhere [2])
- a task in which the driver was directed to “turn around and go back” but not told what type of manoeuvre (U-turn, 3-point turn, drive around the block) should be used to achieve that result
- a task in which the driver was directed to make a turn that required a preparatory lane change but was not told when to make the lane change.

Other new assessment items introduced in Phase 2 of trialling included *Gear Selection* and *Following Distance* (both assessed on 0/1 scales) and *Response to Hazards* (assessed on a 0/1/2 scale).

Feedback from the LTOs involved in trialling indicated that there was a need for a convenient means of recording infrequently-occurring illegal or dangerous driving actions. As a result, four new error codes were shown on the score sheet for each assessable task: *FTS* (fail to stop), *ESL* (exceed speed limit), *NGW* (not give way) and *Other* (any other dangerous or illegal action). These new error categories, plus *Stall* and *Intervention* which were retained from Phase 1 of trialling, were recorded as Y (yes) if the error was committed or were left blank otherwise.

Phase 3

All assessment items that required the LTO to record a rating based on subjective judgement rather than objective criteria were excluded the final phase of trialling. Not only were such items found to be ineffective at discriminating between driver groups [3], but they were also unpopular with the LTOs involved in the trials, who expressed a clear preference for objective criteria.

Feedback from the LTOs involved in earlier phases of trialling also indicated a preference for items to be scored on a simple Yes/No scale, rather than allowing partial credit for partial compliance. This is likely to be a result of the extra workload imposed on LTOs when asked to score on a multi-point scale. As a result, assessment items in the third phase of trialling were simply scored as Yes if all relevant criteria were met or No if one or more criteria were not met. Some items could also be scored as NA if they were not assessable in the circumstances. For example, *Gap Selection* was scored as NA if there was no conflicting traffic present and the applicant did not have to select a safe gap.

The range of error categories recorded ‘by exception’ (that is, marked on the score sheet only if the error was committed) was greatly expanded in Phase 3. Newly defined error categories included *Too Slow*, *Block Pedestrian Crosswalk*, *Mount Kerb* and *Disobey Direction*, among others.

Final assessment framework

The process of selecting driving tasks and assessment items for inclusion in the Drive Test based on their psychometric performance during trialling has been described elsewhere [3]. The resultant assessment framework combines two forms of assessment:

- Credit is awarded, via task assessment items and stage assessment items, for correct performance of specified safe driving skills during the test drive. For example, when making a turn at an intersection, the applicant can receive credit for correct observation, for correct signalling and for correct gap selection.
- Penalties are recorded if the applicant commits a serious error during the test drive – generally a driving action that either is illegal or puts any road user in immediate danger.

Credit items are of two types:

- Task assessment items are assessed only during the performance of specified assessable tasks on the test route. To avoid overloading LTOs, no more than four task assessment items are assessed on any driving task.

- Stage assessment items are assessed on the basis of driving performance across the whole of the first stage or the whole of the second stage of the test route, not only during the specified assessable tasks but also during the additional manoeuvres that are used to link the assessable tasks into a continuous driving route.

Penalty items also belong to two types:

- An Immediate Termination Error is recorded if the applicant commits a driving action that puts any road user in immediate danger. An example is stopping so far past a stop line that the applicant's vehicle projects into the intersecting traffic stream and forces other road users (either drivers or pedestrians) to take evasive action. As the name implies, the test drive is immediately terminated and applicant is deemed to have failed if an Immediate Termination Error is committed.
- A Critical Error is recorded if the applicant commits a serious error that does not put any road user in immediate danger. For example, while entering or leaving a parking space one wheel mounts the kerb but no other road user is affected. The test can continue after a Critical Error is recorded. However, the test is immediately terminated (and the applicant fails) if more than one Critical Error is recorded in Stage 1 or more than two are committed during the entire test.

Immediate Termination Errors and Critical Errors, like stage assessment items, are assessed continuously, both during assessable tasks and linking manoeuvres.

To pass the Drive Test, the applicant must complete the test drive (i.e. not have the test terminated before completion due to an Immediate Termination Error or due to exceeding the threshold number of Critical Errors) and must achieve a specified ratio of items scored Yes to items scored No across all task assessment items and stage assessment items in the test. Items scored as NA (Not Assessable) make no contribution to the test score and have no influence on passing or failing the test.

Final assessment criteria

In defining the assessment criteria for the various assessment items and error categories in the Drive Test, the development team aimed to achieve reliability of assessment through clear specification of objectively verifiable behavioural requirements.

For each task assessment item and each stage assessment item, the criteria list the behaviours that must be observed for the item to be scored as a Yes. The behavioural requirements vary between traffic situations: for example, the requirements for Observation when diverging or changing lanes differ from the requirements when turning left or right at an intersection. Task assessment items are scored as Yes if all behavioural requirements applicable to the traffic situation are satisfied, and No if any applicable requirement is not satisfied. Stage assessment items are scored as Yes if there are not more than two breaches of the requirements throughout the entire stage being assessed, and No if there are more than two breaches.

For each Critical Error and each Immediate Termination Error, the criteria list the driving behaviours that would lead the LTO to record that error type. For some error types, exceptions are also listed. For example, a Critical Error is recorded if the applicant stops the vehicle at a position that intrudes onto or blocks a marked pedestrian crossing; however, an exception is permitted if it is necessary to stop on the crossing to view approaching traffic before completing a turn at an intersection and no pedestrian is affected.

In general, a Critical Error or Immediate Termination Error may be recorded only if the applicant commits one of the driving actions listed in the criteria. However, there are two exceptions. It was not possible for the development team to foresee all the possible errors that might be committed in the course of a test drive; nor would it be desirable to attempt to exhaustively list every possibility in the criteria manual that LTOs must learn and follow. Thus it was necessary to create two error categories that are less precisely defined than the rest. The Immediate Termination Error *Other Dangerous Action* is recorded if the applicant performs a driving action not specified under the other Immediate Termination Errors that results in another road user taking evasive action, or results in danger to road users or

property. The Critical Error *Other Illegal Action* is recorded if the applicant performs an illegal driving action not specified under any other Critical Error that does not lead to another road user taking evasive action or result in danger to road users or property. In either case, the LTO is required to make a brief note of the nature of driving action that resulted in the error being recorded.

The manual used by the LTOs during training and licence testing includes extensive cross-referencing between assessment items and the corresponding error categories, helping to ensure that each driving action is scored in the most appropriate way by the LTO. For example, the criteria for *Gap Selection* state that the applicant is required to select the first safe gap and to reject any unsafe gaps. However, if the applicant's gap selection is so unsafe that it requires another road user to take evasive action, the LTO is referred to the Immediate Termination Error *Fail to Give Way*.

The specification of distances in the assessment criteria has been avoided wherever possible because the LTO, seated in the vehicle, has no means of measuring distances or confirming visual estimates. It was impossible to avoid stating the distance in metres that may be used for the reverse parallel park, but even this is regarded as undesirable, since the accuracy of distance estimates may vary between LTOs.

Simple diagrams have been provided in the LTO manual to illustrate the traffic situations and driving actions described. At present there are 16 diagrams illustrating assessable driving tasks and a further seven illustrating assessment criteria. It is likely that further diagrams will be added in future revisions of the manual. The manual also includes several dozen assessment examples, each giving a description of a traffic situation, the applicant's driving action and the correct assessment of that action based on the criteria. These examples are expected to help LTOs to interpret and apply the assessment criteria correctly.

Concluding comments

Terminating the test drive if the applicant commits an error that puts any road user in immediate danger or requires a road user to take evasive action is an important part of maintaining the safety of applicants, accompanying instructors, LTOs and the travelling public during licence testing. An applicant who commits such an error not only is unready for a solo licence, but is at risk of creating a yet more dangerous situation that other road users may be unable to evade successfully if the test drive were allowed to continue.

Critical Errors are new to licence testing in Victoria, not having been used in the former POLA test. They provide an intermediate level of penalty that is more severe than the loss of a point for an assessment item but less severe than immediate termination of the test drive. Whilst no immediate danger has been created and the test is permitted to continue after a single Critical Error, the commission of a series of such errors indicates that a more dangerous error could be literally "just around the corner" and the drive should be terminated.

The changes made to the assessment framework following the various trials took into account not only the psychometric performance of the items trialled but also the capacity of LTOs to apply complex scoring criteria in a moving vehicle under time pressure. In addition to the three major trials described in this paper, a number of "mini-trials" using only five to twenty learner drivers were also undertaken. The results of the trials determined the number and type of items that could be assessed for each driving manoeuvre. A variety of item scoring protocols were devised and trialled. The eventual decision to score all Drive Test assessment items as a simple Yes or No (plus the possibility of NA for some items) not only eased the assessment workload for the LTOs but also permitted the design of a simpler and more compact score sheet on which to record assessment outcomes. Since the order of driving tasks and assessment items differs between test routes, it would have been difficult to design a single score sheet to be used for all test routes if the various assessment items had required differing layouts for the recording of the assessed score.

The development and refinement of the criteria for the various assessment items and error categories was a cooperative and consultative process, driven initially by input from the development team and feedback from the LTOs involved in the trials, but also involving consultation with other stakeholders within VicRoads and with Victoria Police.

Since the implementation of the Drive Test on 1 July this year, VicRoads has monitored the performance of the new test and of the LTOs who administer it to licence applicants. Further refinements of the assessment criteria will be undertaken as and when required to maximise the reliability of the test by ensuring consistent interpretation and application of the criteria across all licence testing centres in Victoria.

References

1. Cavallo A & Oh S, 'Developing Victoria's new drive test: The GLS context and project management model', *Proceedings, 2008 Australasian Road Safety Research, Policing and Education Conference, 10-12 November 2008, Adelaide, South Australia.*
2. Macdonald WA & Harrison WA, 'Construct validity of Victoria's new licence test', *Proceedings, 2008 Australasian Road Safety Research, Policing and Education Conference, 10-12 November 2008, Adelaide, South Australia.*
3. Harrison WA, 'Getting Rasch about assessment: Using a psychometric approach to select and confirm items in the new Victorian licence test', *Proceedings, 2008 Australasian Road Safety Research, Policing and Education Conference, 10-12 November 2008, Adelaide, South Australia.*
4. Christie R, Cavallo A, Oh S & Johnston D, 'Victoria's new Drive Test: From the drawing board to operational test', *Proceedings, 2008 Australasian Road Safety Research, Policing and Education Conference, 10-12 November 2008, Adelaide, South Australia.*