PUBLISHED PROJECT REPORT PPR062

INTER-RATER RELIABILITY OF THE STROKE DRIVERS SCREENING ASSESSMENT

Version: Final

by J Sentinella, B Sexton and CM Inwood (TRL)

Prepared for: Project Record: PPAD 9/72/62 Validity and Reliability of the Stroke Drivers’ Screening Assessment

Client: Mobility and Inclusion Unit, Department for Transport (Yvonne Brown)

Copyright TRL July 2005

This report has been prepared for Department for Transport, Mobility and Inclusion Unit, is unpublished and should not be referred to in any other document or publication without the permission of Department for Transport. The views expressed are those of the author and not necessarily those of Department for Transport.

<table>
<thead>
<tr>
<th>Approvals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Manager</td>
</tr>
<tr>
<td>Quality Reviewed</td>
</tr>
</tbody>
</table>
This report has been produced by TRL Limited, under/as part of a Contract placed by Department for Transport. Any views expressed are not necessarily those of Department for Transport.

TRL is committed to optimising energy efficiency, reducing waste and promoting recycling and re-use. In support of these environmental goals, this report has been printed on recycled paper, comprising 100% post-consumer waste, manufactured using a TCF (totally chlorine free) process.
CONTENTS

Executive summary i

1 Introduction 3

2 Method 4
   2.1 Recruitment of raters 4
   2.2 Study procedure 4
      2.2.1 Study materials 4
   2.3 Description of the SDSA 5

3 Findings 9
   3.1 Inter-rater reliability 9
   3.2 Raters’ interpretation of scores 10
      3.2.1 Raters’ recommendations 11
   3.3 Errors recorded on the score sheets 12
   3.4 Agreement between an ‘expert’ user and the raters 13
      3.4.1 SDSA scores for Client 1 13
      3.4.2 SDSA scores for Client 2 14
      3.4.3 SDSA scores for Client 3 15
   3.5 Scoring procedure 16
      3.5.1 Observed differences in method of administration and scoring 16
      3.5.2 Difficulties scoring the SDSA 18
      3.5.3 Other comments 19

4 Summary and conclusions 19
   4.1 Inter-rater reliability and interpretation of SDSA scores 19
   4.2 Observed differences in the administration of the SDSA 22
   4.3 Revisions to the SDSA instruction manual 23

5 Acknowledgements 24

6 References 24

Appendix B: Questionnaire 28

Appendix C: Score sheet 32
Executive summary

A stroke can cause physical and cognitive impairments such as lack of concentration, visual inattention and reasoning difficulties which may affect a drivers’ ability to drive safely. The Stroke Drivers Screening Assessment (SDSA) has been developed by the University of Nottingham as a screening tool to identify drivers who are not fit to resume driving following a stroke. As part of a project investigating the validity and reliability of the SDSA TRL and the University of Leeds assessed the inter-rater reliability of the SDSA.

Twenty-eight current users of the SDSA participated in the study. Each rater observed a video of three people completing the SDSA and scored each person’s performance. The raters also completed a questionnaire which asked what result and recommendations they would offer each client, how they themselves administered the SDSA and difficulties they had with scoring and interpretation.

The results showed that the Intra-class correlation coefficients were greater than 0.8 for all components of the SDSA and for the scores, which suggests that there is very high agreement between raters. There were also good levels of agreement between the expert who administered the tasks in the video and the raters who watched the video. The mean rater scores and the expert’s scores were very similar for all components of the SDSA and the standard deviations of the raters’ scores were generally small.

However, the result assigned to each client based upon the pass and fail equations supplied with the SDSA was mixed, especially for clients who were close to the pass-fail borderline. Some raters used a ‘borderline category’ despite no clear definition of a borderline category in the SDSA instruction manual. Other raters felt the calculated result did not reflect the client’s performance in the tasks.

Most raters would have recommended further physical, cognitive and/or driving assessment to drivers who had failed the assessment rather than inform them they were unfit to drive solely on the basis of the SDSA.

A number of calculation errors were identified on the score sheets but in the majority of cases this did not affect the overall result.

Raters reported a number of differences between their own administration of the SDSA and that depicted in the video. A number of raters provided additional information or instruction, especially if the client clearly did not understand the instruction. The SDSA has standard instructions and should be administered exactly as described in the manual. Rephrasing the instructions or providing additional information
could affect the client’s performance in the task and predictive validity of the tests.

Other differences such as providing non-specific feedback and the positioning of the road sign cards in the road sign recognition test were also highlighted.

The difficulties raters found when scoring the test repeated many of those reported in an earlier survey of users. They included scoring the dot cancellation test without a template and scoring the square matrices tests and road sign recognition test without details of the correct answers or card positions in the manual.

The study showed that more guidance is required on how to interpret the results when the overall result is not clear-cut. The possible need for and definition of a ‘borderline’ category should be clarified.

The study’s recommendations include modifying the instruction manual to:

- emphasise the need for standardised administration;
- clarify the positioning of the road sign cards in the road sign recognition test;
- provide advice on non-specific feedback to the client during the assessment;
- facilitate scoring; and
- assist the assessor in interpreting the results for clients who are close to the pass-fail borderline.
1 Introduction

Drivers recovering from a stroke may have both physical and cognitive impairments which could affect their ability to drive safely. These impairments include slowed reactions, perceptual disorders, visual inattention, concentration and reasoning difficulties. At present there are no standard assessment procedures to assess whether drivers who have sustained a stroke can resume driving safely. Some drivers are referred to an accredited Mobility Centre for a driving assessment, which includes an assessment of the person’s physical and cognitive abilities thought relevant to safe driving. However, the large number of people holding driving licences who sustain a stroke make a full driving assessment at a Mobility Centre for each person impractical. It is therefore highly desirable to have a pre-screening tool to help identify those drivers who are not fit to drive – i.e. to identify drivers who are extremely unlikely to pass the full driving assessment.

The Stroke Drivers Screening Assessment (SDSA) was developed by researchers at Nottingham University as a tool to assess fitness to drive. Nouri and Lincoln (1993) found that the SDSA is reasonably successful in predicting road test performance. The SDSA toolkit is sold commercially by Nottingham Rehab Ltd and is used by a number of healthcare professionals, mostly occupational therapists (OTs), to assess fitness to drive.

TRL and the University of Leeds have been commissioned by the Department for Transport to investigate further the reliability and validity of the SDSA. This report summarises a study assessing the inter-rater reliability of the SDSA, which has been undertaken as part of the project.

Inter-rater reliability is concerned with the level of agreement between raters who assess the same clients. If decisions on fitness to drive are to be based upon the results of a screening tool such as the SDSA it is important that the outcomes reached are consistent and dependable.

The objectives of the study were to measure the inter-rater reliability of the SDSA by investigating whether:

- the same result is reached if the test is applied by more than one assessor;
- the scoring procedure of the SDSA is accurate and rigorous; and
- interpretation of the results of the SDSA is accurate and rigorous.

The report is presented in four main sections: the next section outlines the methodology; section 3 describes the results; and section 4 presents the summary and conclusions of the study.
2 Method

2.1 Recruitment of raters

A number of current SDSA users were identified from a recent survey of SDSA users (see Sentinella, Stothart and Keating, unpublished, for more details). These users had indicated that they were willing to take part in further research and supplied contact details. Each SDSA user was contacted by telephone and invited to take part in the study. The users were also invited to distribute additional copies of the study materials to colleagues who currently use the SDSA. All users participating in the study were informed that the data and information supplied by them would be confidential and anonymous.

2.2 Study procedure

Raters were shown a video of three stroke clients using the SDSA, and asked to score each client using the standard SDSA procedure. The raters were also asked to complete a questionnaire. The process was expected to take approximately one hour to complete. A £10 gift voucher was given to those who returned their materials to show appreciation for their help with the study.

Reminder letters were sent to those who had not returned their study materials after 8 weeks. A total of 65 sets of study materials were distributed to SDSA users; 28 were returned.

2.2.1 Study materials

The study materials distributed to current SDSA users included a video cassette, questionnaire, instruction sheet, photocopies of dot cancellation sheets completed by clients 1, 2, and 3, three score sheets (for clients 1, 2, and 3) and two freepost envelopes (one for return of the video, one for the score sheets and questionnaire).

2.2.1.1 Video

The video was specially produced for this study using a professional studio at the television department of the University of Leeds. It depicted the SDSA being administered to three stroke patients by Professor Nadina Lincoln (a developer of the SDSA). The stroke patients were recruited from in-patients at the Chapel Allerton Hospital in Leeds and had given informed consent to take part. A standardised script was
developed to ensure that the filming of the administration of the SDSA to patients was consistent and that the cameraman knew which aspects to film at different points.

2.2.1.2 Instruction sheet
The instruction sheet provided information on how to complete the task, such as pausing the tape to score the square matrices and road sign recognition task. A copy of the instruction sheet can be found in Appendix A.

2.2.1.3 Questionnaire
The questionnaire asked the SDSA users to:

- report problems encountered with the scoring or interpretation of each test;
- compare the type of problems encountered during the study with their own use of the SDSA;
- compare the method of administration of the SDSA shown on the video with their own;
- disclose what result they would award client 1, 2 and 3; and
- provide their recommendations for client 1, 2 and 3 based on the result obtained.

A copy of the questionnaire can be found in Appendix B.

2.3 Description of the SDSA
The SDSA has four component tests:

- dot cancellation test
- square matrix direction test
- square matrix compass test
- road sign recognition test

Dot cancellation test
The dot cancellation test is designed to measure visual scanning, attention and concentration. Clients are asked to cross out all 4-dot groups on a page containing rows of 3-dot, 4-dot and 5-dot groups. They are told that they will be timed during this task, but not that the maximum
time allowed is 15 minutes. Time taken to complete the test is recorded by the assessor, the number of false positives (incorrectly crossed out groups of 3-dot and 5-dot groups) and errors (4-dot groups not crossed out). The time taken to complete the task, and the number of false positives, are used in the final pass and fail equations. The number of errors is not used in these equations.

**Square matrix directions test**

The square matrix directions test is not included in the final pass or fail equations but is included as a ‘practice test’. It is a reasoning task which has four large direction arrows and four small direction arrows. The large arrows correspond to the direction of a ‘lorry’ and the small arrows correspond to the direction of a ‘car’. The arrow direction cards are placed along two sides of a 16 square matrix in front of the client. Cards displaying a car and a lorry facing different directions are placed in front of the client in a pile. The client is asked to position these cards onto the matrix in the correct position. One point is awarded for each correctly placed vehicle, i.e. a maximum of two per card, the maximum score being 32. If more than one card is placed in the same square of the matrix, each card is scored if correctly placed (although this is not explicitly stated in the instructions). The time limit for this test is 5 minutes. Clients are told to stop the task when the time limit is reached using a statement such as “that’s fine, you can stop now”.

![Figure 1: Square matrix directions test layout](Photo: Lincoln et al., 2004)
Square matrix compass test

The square matrix compass test uses the 16 square matrix with ‘compass cards’ replacing the arrow cards. These cards depict arms of a compass, the black arm indicating the direction of travel. This time the cards placed in front of the client display two vehicles travelling in different directions away from the centre of a roundabout. The clients are asked to place the card in the square that matches the direction of the cars with the compass arrows. They are told that there are more cards than available spaces and that therefore some cards do not fit in the matrix. The time limit for this test is 5 minutes and it is scored in the same way as the square matrix direction test. Clients are told to stop the task when the time limit is reached using a statement such as “that’s fine, you can stop now”.

![Figure 2: Square matrices compass test layout](Photo: Lincoln et al., 2004)

Road sign recognition test

In the road sign recognition test, the client is asked to match two sets of cards for this test. One set (of 20 cards) shows various road signs, and the other set (of 12 cards) shows road traffic situations. On each road traffic situation card, the shape of the missing road sign is evident. Clients are asked to place the road sign card on top of the road situation card which they think it matches best, based on the information available in the scene. The diagram in the manual illustrates that the road situation cards should be placed in front of the client spaced out in a four
by three arrangement, with example card and the road sign cards in a pile in front. Each correct match scores one point, with a maximum of 12 points. When more than one road sign card is placed on a road situation card, if one is correct it is still scored. The client is not informed that the time limit for this test is 3 minutes or that there are more road sign cards than situation cards. Clients are told to stop the task when the time limit is reached using a statement such as “that’s fine, you can stop now”.

**Scoring**

With the exception of the dot cancellation task errors and false positives calculations which are calculated after the SDSA test has been completed, each task of the SDSA is scored before the client commences the next task. The scores recorded include:

A: time in seconds taken to complete the dot cancellation task

B: errors in the dot cancellation task

C: false positives in the dot cancellation task

D: square matrix directions score

E: square matrix compass score

F: road sign recognition score

Scores for A, C, E and F are then put into a pass equation and a fail equation to produce two scores, the higher of which indicates whether the client has ‘passed’ or ‘failed’. The following equations are used:

Pass score = \( (A \times 0.012) + (C \times 0.216) + (E \times 0.409) + (F \times 1.168) - 13.79 \)

Fail score = \( (A \times 0.017) + (C \times 0.035) + (E \times 0.185) + (F \times 0.813) - 10.042 \)

The SDSA instruction manual states that if the client has a higher value for the pass equation than for the fail equation (referred to as a ‘pass’ result in this report) his or her “cognitive abilities are such that driving is feasible but physical and mental aspects will need to be checked by a medical practitioner or through a Disabled Driving Centre [Mobility Centre]” (Nouri and Lincoln, undated). If the value of the fail equation is higher (a ‘fail’ result) the client should be advised not to drive but may be reassessed in 3-4 months, to minimise practice effects. The manual suggests that clients who are ‘borderline’ are referred to a Mobility Centre for further assessment.
3 Findings

3.1 Inter-rater reliability

The analyses were based upon the SDSA scores for 3 clients, each scored by 28 different raters: a total of 84 SDSA scores. Inter-rater reliability was assessed using the intra-class correlation statistic.

After the raters’ calculations for pass and fail scores were corrected (11 out of 84), the intra-class correlation coefficient (ICC) was computed for pass and fail scores and each component of the SDSA. Table 1 illustrates that the ICCs are all high (ranging from 0.82 to 1.0), suggesting a very high level of agreement between raters for each of the components of the SDSA.

<table>
<thead>
<tr>
<th>Component</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Time for dot cancellation (secs)</td>
<td>1.00</td>
</tr>
<tr>
<td>B: Errors in dot cancellation</td>
<td>0.98</td>
</tr>
<tr>
<td>C: False positives in dot cancellation</td>
<td>0.86</td>
</tr>
<tr>
<td>D: Square Matrix Directions score</td>
<td>0.98</td>
</tr>
<tr>
<td>E: Square Matrix Compass score</td>
<td>0.95</td>
</tr>
<tr>
<td>F: Road sign recognition score</td>
<td>0.99</td>
</tr>
<tr>
<td>Pass score</td>
<td>0.82</td>
</tr>
<tr>
<td>Fail score</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Table 2 shows the level of agreement between raters of a ‘pass’ or ‘fail’ result based upon whether the ‘pass’ or ‘fail’ score was greater. It indicates that there is a completely consistent outcome for the first client rated, a mixed outcome for the second, and a fairly consistent one for the third. Investigating further the inconsistency with the second client, it appears that the ‘pass’ and ‘fail’ scores were quite similar so that many of those with a ‘pass’ result only had a slightly higher ‘pass’ score than the ‘fail’ score.
Table 2: Level of pass/fail agreement between raters

<table>
<thead>
<tr>
<th>Pass or fail</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client 1</td>
<td>28</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Client 2</td>
<td>10</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>Client 3</td>
<td>25</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>63</td>
<td>21</td>
<td>84</td>
</tr>
</tbody>
</table>

3.2 Raters’ interpretation of scores

The questionnaire asked raters to indicate what result they would assign each client based upon their own calculations and judgements. Table 3 indicates that a ‘borderline’ result was frequently used. In three cases the rater gave an ‘other’ result because the result the pass and fail scores indicated was felt to be inappropriate (e.g. the client should not have ‘passed’).

Table 3: Result assigned by raters

<table>
<thead>
<tr>
<th></th>
<th>Pass</th>
<th>Fail</th>
<th>Borderline</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client 1</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>Client 2</td>
<td>1</td>
<td>18</td>
<td>8</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>Client 3</td>
<td>16</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>21</td>
<td>18</td>
<td>3</td>
<td>84</td>
</tr>
</tbody>
</table>

As indicated in section 3.1, there was a high level of agreement between raters for the result for client 1 and a mixed outcome for client 2. However, there was a lower level of agreement in the result assigned to client 3 than might be expected from the results shown in Table 2. For both client 2 and client 3 a significant minority of raters considered the result to be ‘borderline’ despite, as one rater commented, there being “no clear criteria on what constitutes a borderline”.

Of the ‘borderline’ results, 15 were ‘passes’ and 3 were ‘fails’. For client 2 the difference between pass and fails scores ranged from -0.28 to +0.53, whilst the difference between the scores for client 3 ranged from -1.11 to +1.03.

Other raters explained that they would not give a ‘borderline’ result:

“There are no guidelines given in scoring manual regarding what constitutes a ‘borderline’ result. Therefore I always give a pass or fail and recommend further assessment (by test centre) if a pass
and re-assessment in 3-4 months if a fail (if [the] client [is] likely to change).”

“The test only allows a pass or a fail, though the difference is 0.7 which is probably more likely to be borderline. I would use my knowledge of the client to make the final decision.”

3.2.1 Raters’ recommendations

Table 4 indicates the recommendations the raters would have made based upon the results of the SDSA. It shows that most raters would have recommended that the client undergo further physical and/or cognitive assessment whatever the result: the SDSA would generally not be used alone to assess fitness to drive. Only one rater recommended that a client be told that they were unfit to drive based upon the SDSA results only. In four cases, the rater gave the client a ‘pass’ or ‘borderline’ result and recommended that they were safe to drive without any further assessment.

The raters’ recommendations did not always follow the guidance given in the SDSA instruction manual. In 14 cases the rater would have told the client they were unsafe to resume driving despite a ‘pass’ or ‘borderline’ result and in three cases the rater would not have advised a client with a ‘fail’ result not to drive. Only three of the 18 clients with a ‘borderline’ result would have been referred to a Mobility Centre for further assessment.

1 This conclusion would not necessarily apply to clients other than the three tested.
Table 4: Raters’ recommendation based upon SDSA result (n=84)

<table>
<thead>
<tr>
<th>SDSA result</th>
<th>Recommendation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Safe to resume driving</td>
</tr>
<tr>
<td>Pass (n=42)</td>
<td>13</td>
</tr>
<tr>
<td>Fail (n=21)</td>
<td>-</td>
</tr>
<tr>
<td>Borderline (n=18)</td>
<td>1</td>
</tr>
<tr>
<td>Other (n=3)</td>
<td>-</td>
</tr>
</tbody>
</table>

* multiple recommendations were given so a recommendation of ‘safe’ or ‘unsafe’ to resume driving may have been given together with a referral for further physical and/or cognitive and/or driving assessment.

In some cases the rater felt the client’s poor performance was not reflected in the SDSA result. Their recommendations were, therefore, based upon their observations of the client performing the tasks. As the following comments illustrate:

“I personally would not let him drive but despite lots of errors in the cancellation test his score is a pass.”

“Objectively based on the score she was borderline, but I would have grave concerns about her ability to drive based on observation of her performance!”

3.3 Errors recorded on the score sheets
Eight raters made 11 errors when working out the pass and fail scores using the equations:
- in ten cases this did not affect the result; and
- in one case a ‘fail’ result was obtained when it would have been a ‘pass’.

In one case the calculation error was due to an incorrect score being copied from the individual scores into the pass or fail equation (Square Matrix Directions score instead of the Square Matrix Compass score).
One rater left the dot cancellation time in minutes rather than seconds for one client, another rater used ‘0’ in the equation instead of the time recorded when the client had run out of time for the task. Other raters made errors when multiplying or summing the SDSA components in the equations.

3.4 Agreement between an ‘expert’ user and the raters

The SDSA was administered in the video by one of the developers of the tool, Professor Nadina Lincoln. The scores she gave each client (referred to as the ‘expert’ score) were compared to the scores assigned by the 28 raters who watched the video.

As highlighted in section 3.3, some of the scores recorded on the score sheets were transcribed incorrectly into the equations and/or the pass or fail scores were calculated incorrectly. For these analyses corrected scores were used to examine how close the agreement was between the expert and raters for each client.

The shaded lines in Tables 5 to 7 indicate the scores not included in the pass or fail equations.

3.4.1 SDSA scores for Client 1

Table 5 illustrates that the raters scored each task similarly to the expert rater: the mean raters’ scores were close to the expert scores. Three-quarters of raters obtained the same score as the expert for the square matrix compass test and 68% of raters agreed with the expert’s score for the road sign recognition test. The scores for the dot cancellation test were less similar to the expert’s scores.
### Table 5: Comparison between ‘expert’ and rater SDSA scores for client 1 (n=28)

<table>
<thead>
<tr>
<th>Component</th>
<th>‘Expert’ score</th>
<th>Raters Mean score</th>
<th>Standard deviation</th>
<th>Min</th>
<th>Max</th>
<th>Mode</th>
<th>% giving same score as expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Dot cancellation time (secs)</td>
<td>204</td>
<td>205.32</td>
<td>2.29</td>
<td>201</td>
<td>213</td>
<td>206</td>
<td>29%</td>
</tr>
<tr>
<td>B: Errors (dot cancellation)</td>
<td>113</td>
<td>111.57</td>
<td>4.12</td>
<td>99</td>
<td>120</td>
<td>112</td>
<td>18%</td>
</tr>
<tr>
<td>C: False positives (dot cancellation)</td>
<td>28</td>
<td>29.25</td>
<td>6.28</td>
<td>17</td>
<td>57</td>
<td>29</td>
<td>11%</td>
</tr>
<tr>
<td>D: Square Matrix Directions score</td>
<td>21</td>
<td>19.96</td>
<td>2.98</td>
<td>10</td>
<td>23</td>
<td>21</td>
<td>61%</td>
</tr>
<tr>
<td>E: Square Matrix Compass score</td>
<td>27</td>
<td>25.46</td>
<td>4.27</td>
<td>10</td>
<td>23</td>
<td>27</td>
<td>75%</td>
</tr>
<tr>
<td>F: Road sign recognition score</td>
<td>9</td>
<td>8.89</td>
<td>0.57</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>68%</td>
</tr>
<tr>
<td>Pass score</td>
<td>16.26</td>
<td>15.79</td>
<td>2.56</td>
<td>8.46</td>
<td>23.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fail score</td>
<td>6.72</td>
<td>6.41</td>
<td>1.03</td>
<td>3.45</td>
<td>8.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Equation (pass–fail score)</td>
<td>9.54</td>
<td>9.38</td>
<td>1.62</td>
<td>5.01</td>
<td>15.10</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

#### 3.4.2 SDSA scores for Client 2

The raters’ scores for client 2 were also very similar to the expert’s scores. The mean score was very close to the expert score and the standard deviation for each task was less than 2 (ranging from 0.48 to 1.93) as shown in Table 6. The majority of raters watching the video gave the client 13 points for the road sign recognition test whilst the expert gave a score of 12.
Table 6: Comparison between ‘expert’ and rater SDSA scores for client 2 (n=28)

<table>
<thead>
<tr>
<th>Component</th>
<th>‘Expert’ score</th>
<th>Raters Mean score</th>
<th>Standard deviation</th>
<th>Min</th>
<th>Max</th>
<th>Mode</th>
<th>% giving same score as expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Dot cancellation time (secs)</td>
<td>900</td>
<td>900</td>
<td>1.36</td>
<td>895</td>
<td>905</td>
<td>900</td>
<td>93%</td>
</tr>
<tr>
<td>B: Errors (dot cancellation)</td>
<td>70</td>
<td>70.75</td>
<td>1.92</td>
<td>67</td>
<td>74</td>
<td>71</td>
<td>11%</td>
</tr>
<tr>
<td>C: False positives (dot cancellation)</td>
<td>26</td>
<td>25.68</td>
<td>1.85</td>
<td>21</td>
<td>26</td>
<td>29</td>
<td>32%</td>
</tr>
<tr>
<td>D: Square Matrix Directions score</td>
<td>6</td>
<td>5.71</td>
<td>1.18</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>75%</td>
</tr>
<tr>
<td>E: Square Matrix Compass score</td>
<td>12</td>
<td>12.46</td>
<td>1.93</td>
<td>6</td>
<td>15</td>
<td>13</td>
<td>7%</td>
</tr>
<tr>
<td>F: Road sign recognition score</td>
<td>1</td>
<td>1.32</td>
<td>0.48</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>68%</td>
</tr>
<tr>
<td>Pass score</td>
<td>8.702</td>
<td>9.2</td>
<td>1.17</td>
<td>5.82</td>
<td>10.71</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fail score</td>
<td>9.207</td>
<td>9.54</td>
<td>0.55</td>
<td>8.02</td>
<td>10.38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Equation (pass–fail score)</td>
<td>-0.505</td>
<td>-0.34</td>
<td>0.66</td>
<td>-2.20</td>
<td>0.53</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

3.4.3 SDSA scores for Client 3

Table 7 indicates high levels of agreement between the raters and the expert for most scores. There was less agreement between the raters and the expert on the time taken for the dot cancellation test and the number of errors for this test (although the latter is not included in the pass or fail equations).
Table 7: Comparison between ‘expert’ and rater SDSA scores for client 3 (n=28)

<table>
<thead>
<tr>
<th>Component</th>
<th>‘Expert’ score</th>
<th>Raters Mean score</th>
<th>Standard deviation</th>
<th>Min</th>
<th>Max</th>
<th>Mode</th>
<th>% giving same score as expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Dot cancellation time (secs)</td>
<td>467</td>
<td>464.61</td>
<td>13.07</td>
<td>409</td>
<td>472</td>
<td>470</td>
<td>7%</td>
</tr>
<tr>
<td>B: Errors (dot cancellation)</td>
<td>43</td>
<td>39.96</td>
<td>8.26</td>
<td>0</td>
<td>49</td>
<td>43</td>
<td>32%</td>
</tr>
<tr>
<td>C: False positives (dot cancellation)</td>
<td>0</td>
<td>1.43</td>
<td>7.37</td>
<td>0</td>
<td>39</td>
<td>0</td>
<td>93%</td>
</tr>
<tr>
<td>D: Square Matrix Directions score</td>
<td>32</td>
<td>30.86</td>
<td>4.2</td>
<td>16</td>
<td>32</td>
<td>32</td>
<td>93%</td>
</tr>
<tr>
<td>E: Square Matrix Compass score</td>
<td>16</td>
<td>15.43</td>
<td>2.36</td>
<td>8</td>
<td>20</td>
<td>16</td>
<td>86%</td>
</tr>
<tr>
<td>F: Road sign recognition score</td>
<td>9</td>
<td>9.25</td>
<td>0.59</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>61%</td>
</tr>
<tr>
<td>Pass score</td>
<td>8.87</td>
<td>9.21</td>
<td>2.19</td>
<td>5.61</td>
<td>18.49</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fail score</td>
<td>8.174</td>
<td>8.28</td>
<td>8.23</td>
<td>6.71</td>
<td>10.39</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Equation (pass –fail score)</td>
<td>0.053</td>
<td>0.93</td>
<td>1.53</td>
<td>-1.11</td>
<td>8.10</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

3.5 Scoring procedure

3.5.1 Observed differences in method of administration and scoring

The SDSA has a standardised method of administration. Raters’ were asked to give details of whether their own administration of the SDSA differed from that depicted in the video. Over a quarter of raters claimed that the method of administration of the SDSA shown in the video was different to their own. Details of how they performed the SDSA differently encompassed:

- verbal instructions/ additional information given
- presentation of cards
- timing of tests
• scoring
• practical issues

3.5.1.1 Verbal instructions/ additional information
Raters reported that they give additional information to clients or re-word the instructions, especially to clients who do not appear to understand what is required of them. Examples of additional information given include:

• Informing the client of the time allocation for all the tests
• Telling the client that there are more sign cards than scenarios in road sign recognition
• Re-wording the instructions for the square matrix test
• Telling the client that if he or she crosses out a group of 3- or 5-dots mistakenly, this can be indicated by ringing them
• Adding “Are you ready? Then start now” prior to each task to help timing

One rater pointed out that in contrast to their own method of administration, the assessor on the video said “you’ve done fine” after two of the clients had completed the SDSA.

3.5.1.2 Presentation of cards in the road sign recognition test
Several raters commented that they present the road sign cards in a pile rather than spread across the table in front of the client as shown in the video.

3.5.1.3 Timing of the tests
The first row of the dot cancellation task is used as a practice and clients are given feedback on any errors they make before commencing with the rest of the task. One rater reported that they previously included the practice line in the timing of the dot cancellation task. However, the manual states that this should not be included.

3.5.1.4 Scoring of the dot cancellation test
One rater reported that if a client realised he or she had made an error on the dot cancellation test, she would not include it in her error/false positive calculations.
3.5.1.5 Practical issues

It was suggested that instead of the rater in the video holding the test paper for one of the clients, it could have been taped to the table.

3.5.2 Difficulties scoring the SDSA

Half of the raters reported difficulties with interpreting or scoring the SDSA tests. The majority of difficulties replicate those reported in the survey of users (Sentinella, Stothart and Keating, unpublished). Raters reported that many of the difficulties they faced scoring the clients on video were typical of those normally encountered in the routine use of the SDSA.

Other difficulties reported were mainly associated with the study materials and the disadvantages of using a video technique to observe clients. These included the difficulty of seeing the cards clearly when the video was paused at the end of each test.

Typical examples of the difficulties encountered for each task are given below.

3.5.2.1 Dot Cancellation test

Six raters highlighted the difficulty they had with the visibility of the markings on the completed dot cancellation sheets; as one rater explained, “the thick pencil obscured some of the dots”.

Raters also reported feeling uneasy over the accuracy of counting the markings, for example:

“It is a nightmare test to score. I am sure I am constantly making mistakes.”

An overlay sheet or template to aid scoring was also a theme running through raters' remarks:

“Scoring is time-consuming. It would benefit from a scoring template, for example on a transparent page.”

“An overlay for dot cancellation test would make scoring more reliable.”

3.5.2.2 Square Matrix tests

Some raters commented that the poor quality of the image when the video was paused made scoring the square matrix directions and
compass tests difficult. One rater explained that she had devised a system to minimise errors, by looking for the “pattern of lorries and then cars” and “the pattern and continuity of compass points”.

3.5.2.3 Road sign recognition test

Raters did not always find the task of matching the road sign cards to the road situation cards an easy one, as one rater explained:

“I always have problems with one card. I do not know if it is ‘no entry’ or ‘no left turn’, I suspect the former.”

3.5.3 Other comments

Concern was expressed that the SDSA does not utilise all available information, such as the errors (component ‘B’) made during the dot cancellation task and the square matrix directions test score (component ‘D’). One rater said that in trying to perform quickly a client could fail to cross out all the correct dots in the dot cancellation task. This would result in fewer false positives being recorded, and this would produce a favoured result because of the form of the pass and fail equations.

4 Summary and conclusions

4.1 Inter-rater reliability and interpretation of SDSA scores

Inter-rater reliability of the SDSA was assessed using the intra-class correlation coefficient (ICC) statistic. The results showed that the ICCs were greater than 0.8 for all components of the SDSA and the pass and fail scores, which suggests that there is very high agreement between raters.

However, the level of agreement between raters for the result that should be assigned to each client was mixed. The SDSA result should be based upon whether the ‘pass’ score or ‘fail’ score is greater. For two clients shown in the video the difference between the pass and fail scores was small, consequently, the result given by raters was mixed. A ‘borderline’ result was frequently used, although there was confusion and difference of opinion over what constituted a ‘borderline’ result or even if it existed. The SDSA instruction manual simply states that clients “who are borderline may be advised to attend a Disabled Driving Centre for further assessment” (Nouri and Lincoln, undated). It does not give details on the size of difference between scores that would constitute a ‘borderline’ result. In this study the difference between pass
and fail equations (pass – fail score) ranged from -1.11 to +1.03 for the ‘borderline’ category. A study undertaken by Lundberg et al. (2003), which validated a Nordic version of the SDSA with 97 stroke patients, has suggested that a borderline category of -0.49 to +0.49 is used.

Concern was expressed by some raters that the result of the SDSA did not reflect the clients’ performance in the tasks, perhaps suggesting a lack of confidence in the SDSA’s ability to predict driving performance. The instruction manual could be modified to include an overview of the sub-tests to assist with understanding and help instil greater confidence in its use, as suggested in the survey of users report (Sentinella et al., unpublished).

It is evident from the findings that a ‘pass’ or ‘fail’ result does not necessarily translate to a recommendation of safe to drive or unsafe to drive respectively. The SDSA instruction manual states that if the client has a higher pass score than fail score (‘pass’ result), this indicates that the client is cognitively able to drive, but should be assessed physically and mentally by a medical practitioner or through a Mobility Centre. If the client has a higher fail score than pass score (‘fail’ result), he or she should be advised not to drive but may be re-assessed after 3-4 months to check whether abilities have improved (Nouri and Lincoln, undated). In this study only one rater recommended that the client be told that he or she was unfit to drive based upon the SDSA results only. The majority of raters recommended that a client should undergo further cognitive, physical or driving assessment, despite the result. Raters do not appear to use the tool as a screening test to identify clients requiring further assessment. Rather, there is a tendency to advise clients to have further assessments whatever the outcome of the SDSA².

The comparison of mean rater scores and ‘expert’ scores showed that the most of the test scores were very similar. Generally, the mean raters’ score was very close to the ‘expert’ score and the standard deviations were small, suggesting good levels of agreement between the expert who administered the tasks in the video and the raters who watched the video.

Some slight variation in scores between the ‘expert’ and raters was anticipated in this study as the raters were scoring the SDSA in a different setting to the expert (i.e. observing the video footage). A number of raters said that they found some tasks difficult to score from the video, for example, the final view of the square matrices tests and road sign recognition test when the video was paused. The time

² As pointed out in section 3.2.1, this conclusion may not apply to clients who perform differently from those tested here.
recorded for the dot cancellation task by raters watching the video was also likely to slightly differ from the expert as it would be difficult to press the stopwatches at exactly the same time.

The differences in scores for errors and false positives in the dot cancellation test may also reflect the difficulties raters reported scoring the test. Raters were provided with copies of the completed dot cancellation sheet to score: some raters said the thick pen used by the clients in this study obscured some of dots making it difficult to score. Other raters found the dot cancellation task difficult to score without a template.

Eleven of the 84 SDSA score sheets contained calculation errors made by the raters. The majority of errors were made when multiplying or summing the SDSA components in the equations. Other errors included recording the time for the dot cancellation test in minutes rather than seconds and transcribing an incorrect figure into an equation from the individual scores. In the majority of cases the errors did not affect the overall result (i.e. whether the pass or fail equation was greater).

It is not clear why the scores for dot cancellation test errors (B) and the square matrix directions test (D) are recorded on the score sheet when they are not used in the pass or fail equations. The SDSA instruction manual does not include any advice on interpreting individual scores and the inclusion of the scores increases the chance of making an error in transcribing individual scores into the equations. The dot cancellation test is also time consuming to score.

There is no indication from this study that raters’ used individual scores in their interpretation of the result, although the questionnaire did not seek this information directly.

Despite the difficulties encountered by raters scoring the tests, the findings indicate that the SDSA has good inter-rater reliability. However, more guidance is required on how to interpret the results when the scores for the pass and fail equations are close. The use of a ‘borderline’ category should be clarified and, if used, clearly defined.

The accuracy of the scoring procedure may be improved by:

- providing a template/overlay to assist scoring the dot cancellation test;
- providing a full set of the correct answers to the square matrices and road sign recognition tests (perhaps labelling the back of the cards with the corresponding matrix location/road situation);
- redesigning the score sheet so it excludes scores that are not used and so minimise errors; and/or
• providing a spreadsheet or an electronic version of the score sheet incorporating the formulae to calculate the pass and fail equations correctly.

4.2 Observed differences in the administration of the SDSA

All the 28 raters who took part in the study were current users of the SDSA. Eleven raters reported that their own administration of the SDSA differed from that shown in the video. Some raters said they gave additional instructions or information to the client, especially if the client did not appear to understand the instructions given in the SDSA instruction manual.

Another common difference in administration was in the presentation of the cards in the road sign recognition test. Some raters commented that they placed the road sign cards in a pile in front of the client, rather than spread across the table in front of the client as shown in the video. If cards are spread across the table, clients may find it easier to select the matching sign for the road situation cards by looking at the situation card first and then the signs. If the cards are presented in a pile, this may lead the client to look at the road sign card and try to place it in the correct place, potentially ‘wasting time’ as clients are not warned that there are more sign cards than situations and the test is time limited. Some raters warn clients that there are more sign cards than situations and the time allocated for each test. The placement of cards needs to be clarified in the instruction manual.

When scoring the dot cancellation test one rater reported that she previously included the practice line in the timing of the dot cancellation task. However, the manual states that this should not be included.

Each test should be administered exactly as described in the manual and instructions read out verbatim. The SDSA was validated using the instructions given in the instruction manual: rephrasing the instructions or providing additional information could affect the client’s performance in the task and predictive validity of the tests. The instruction manual should include information on how to deal with clients who are unable to understand the instructions.

One rater noted that the assessor commented at the end of session, “you’ve done fine”. As the assessments were set-up for this study, the clients shown in the video were not told their SDSA results or given recommendations based upon their performance. Telling the client that he or she had “done fine” could mislead him or her, especially if the client was subsequently told he or she had failed the SDSA. The assessor did use non-specific feedback during the assessments shown
on video, for example: “Ok, I’m going to stop you there, that’s fine for
that one”. This was to encourage the client to continue with the next
task. However, the rater’s comment suggests that the type and level of
feedback that is permitted should be clarified in the instruction manual.

4.3 Revisions to the SDSA instruction manual

The SDSA instruction manual quoted in this report is the manual that is
included in the SDSA toolkit and used by most SDSA users (Nouri and
Lincoln, undated). A new revised instruction manual is now available
from the toolkit developers on request (Lincoln et al., 2004). The revised
instruction manual includes a number of the recommendations described
above such as:

- Providing advice on how to deal with clients who are unable to
  understand the instructions;
- Clarification of how the road sign cards should be placed in the
  road sign recognition test;
- Correct answers to the road sign recognition test;
- Use of non-specific feedback such as ‘that’s fine’ to encourage
  clients to persevere;
- Brief description of the content validity of the test.

Although some raters reported that the use of a felt-tipped pen made the
dot cancellation test difficult to score, the revised manual recommends
the use of this type of pen because it is easier for clients to use than a
pencil or biro.

The revised instruction manual has slightly altered its recommendations
for clients who fail the assessment. It now suggests that individual
scores, including the dot cancellation errors and square matrix directions
test, are compared with the scores of criterion groups who passed or
failed the road test to provide additional information to aid interpretation
of why a client has failed the assessment. However, the dot cancellation
errors score and square matrix directions test are not included in the
SDSA pass and fail equations and their content validity in relation to on
road driving has not been clearly established. It is not entirely clear,
therefore, how useful the information from these scores would be in
practice.

The advice to refer ‘borderline’ clients to a driving mobility centre has
been excluded from the revised manual. Although ‘borderline’ results
are discussed in relation to other studies, a ‘borderline’ category is not
clearly defined.
The following recommendations are not currently covered in the revised manual and could be considered:

- Advice on interpreting the pass and fail equations when they are close, including guidance on the use and definition of a ‘borderline’ category;
- Provide a template/overlay to assist scoring the dot cancellation task;
- Provide a full set of correct answers to the square matrices tests;
- Redesign the score sheet to exclude scores which are not used in the pass and fail equations;
- Make an electronic score sheet or spreadsheet available, which automatically calculates the pass and fail equations.

5 Acknowledgements

The work described in this report was carried out in the Safety Group of TRL Limited and the Institute of Transport Studies, the University of Leeds. The authors are grateful to Professor Nadina Lincoln, University of Nottingham for administering the tests and the clients who participated. They also wish to thank Chris Baughan, TRL’s Technical Referee for this study, who carried out the quality review and auditing of this report, to Dr Liliana Read, University of Leeds, who helped develop the methodology, and to those who gave their time to score the video.

6 References


Appendix A: Instruction sheet

SDSA INTER-RATER RELIABILITY
INSTRUCTIONS

Please follow the instructions given below for each of the patients shown in the video. You will need to use the materials we have provided to complete this task (scoring sheets, dot cancellation outputs, and questionnaire) and you will also need a stopwatch to monitor the time it takes the patients to complete the tasks. The scoring sheets and dot cancellation outputs are numbered according to the patient they refer to in the video (1, 2 or 3 based on their order in the video) so that the information can be linked.

Dot Cancellation Task
- Start timing when the patient begins the dot cancellation task
- Finish timing when the patient completes the task or the time limit has expired
- Record the time taken to complete the task on the appropriate score sheet
- Pause the tape
- Use the dot cancellation output to evaluate the patients performance
- Record the results for this task on the appropriate score sheet
- Resume playing the tape

Square Matrix – Directions Task
- Start timing when the patient begins the square matrix task
- Finish timing when the patient completes the task or the time limit has expired
- Pause the video when the completed square matrix is displayed and evaluate the patients performance
- Record the results for this task on the appropriate score sheet
- Resume playing the tape

Square Matrix – Compass Task
- Start timing when the patient begins the square matrix task
- Finish timing when the patient completes the task or the time limit has expired
• Pause the video when the completed square matrix is displayed and evaluate the patients performance
• Record the results for this task on the appropriate score sheet
• Resume playing the tape

**Road Sign Recognition Task**

• Start timing when the patient begins the square matrix task
• Finish timing when the patient completes the task or the time limit has expired
• Evaluate the patients performance as each pair of road sign and road situation cards are presented
• Record the results for this task on the appropriate score sheet

Please repeat all instructions for all patients seen on the video. When all 3 score sheets have been completed please fill out the questionnaire provided.
Appendix B: Questionnaire

Please complete this questionnaire by ticking the appropriate boxes and filling in the spaces as required. Any information you provide will be treated in the strictest confidence and used for research purposes only.

**Q1 Did you encounter any problems in interpreting or scoring the dot cancellation task for any of the clients?**

*(Please tick one box only)*

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

⇒ If “Yes” please give details below:

**Q2 Did you encounter any problems in interpreting or scoring the square matrix direction test for any of the clients?**

*(Please tick one box only)*

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

⇒ If “Yes” please give details below:

**Q3 Did you encounter any problems in interpreting or scoring the square matrix compass test for any of the clients?**

*(Please tick one box only)*

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

⇒ If “Yes” please give details below:
Q4  Were these problems typical of those you normally encounter in your routine use of the SDSA?
(Please tick one box only)
Yes  ☐
No  ☐  ⇒ If “No” please give details below:
Not applicable (no problems reported)

Q5  Was the method of administration of the SDSA shown in the video the same as your own?
(Please tick one box only)
Yes  ☐
No  ☐  ⇒ If “No” please give details below:

Q6  What result did you give client 1?
(Please tick one box only)
Pass  ☐
Fail  ☐
Borderline  ☐
Other  ☐  Please give details below:

Q7  What result did you give client 2?
(Please tick one box only)
Pass  ☐
Fail  ☐
Borderline  ☐
Other  ☐  Please give details below:
Q8 What result did you give client 3?
(Please tick one box only)
Pass 
Fail 
Borderline 
Other 
Please give details below:

Q9 What recommendations would you make, concerning client 1’s fitness to drive, based on this result?
(Please tick all that apply)
Safe to return to driving 
Unsafe to return to driving 
Further cognitive assessment required 
Physical assessment needed 
Other 
Please give details below:

Q10 What recommendations would you make, concerning client 2’s fitness to drive, based on this result?
(Please tick all that apply)
Safe to return to driving 
Unsafe to return to driving 
Further cognitive assessment required 
Physical assessment needed 
Other 
Please give details below:
Q11 What recommendations would you make, concerning client 3’s fitness to drive, based on this result?

(Please tick all that apply)

- Safe to return to driving
- Unsafe to return to driving
- Further cognitive assessment required
- Physical assessment needed
- Other

Please give details below:

Thank you very much for your help with this task.

Please return the SDSA score sheets and the questionnaire to TRL in the pre-paid envelope provided. Alternatively, if you do not have a pre-paid envelope please address the test materials to:

Catherine Inwood
TRL Limited
FREEPOST (Licence no SCE1560)
Old Wokingham Road
Crowthorne
Berkshire
RG45 6AU
# Appendix C: Score sheet

## DRIVING ASSESSMENT SCORE SHEET

<table>
<thead>
<tr>
<th>NAME</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADDRES</td>
<td>ASSESSOR</td>
</tr>
</tbody>
</table>

Enter all scores into the appropriate sections of the table below.

<table>
<thead>
<tr>
<th>TASK</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square Matrix</td>
<td>Directions Score [D] ______</td>
</tr>
<tr>
<td>Road Sign Recognition</td>
<td>Score [F] ______</td>
</tr>
</tbody>
</table>

Transfer the relevant scores into both equations below

### PASS EQUATION

\[
\begin{align*}
  [A] \quad & \times 0.012 = \quad \_ \_ \_ \\
  [C] \quad & \times 0.216 = + \quad \_ \_ \_ \\
  [E] \quad & \times 0.409 = + \quad \_ \_ \_ \\
  [F] \quad & \times 1.168 = + \quad \_ \_ \_ \\
  \quad & \quad \_ \_ \_ \\
  (\text{Constant}) & - 13.79 \\
\text{Total} & = \quad \_ \_ \_ \\
\end{align*}
\]

### FAIL EQUATION

\[
\begin{align*}
  [A] \quad & \times 0.017 = \quad \_ \_ \_ \\
  [C] \quad & \times 0.035 = + \quad \_ \_ \_ \\
  [E] \quad & \times 0.185 = + \quad \_ \_ \_ \\
  [F] \quad & \times 0.813 = + \quad \_ \_ \_ \\
  \quad & \quad \_ \_ \_ \\
  (\text{Constant}) & - 10.042 \\
\text{Total} & = \quad \_ \_ \_ \\
\end{align*}
\]