Predicting fitness to drive in people with physical and/or cognitive impairment using a clinical test

Lynnette G. Kay, Anita C. Bundy, & Lindy M. Clemson

Occupational Therapy,

Faculty of Health Sciences,

The University of Sydney

Address all correspondence to Anita Bundy, PO Box 170 Lidcombe NSW 1825 Australia,

Telephone: 61 2 9351 9373, Fax: 61 2 9351 9346, Email: a.bundy@usyd.edu.au

Keywords: automobile driving, rehabilitation, Rasch analysis,

Abstract

Objective. To examine the construct and predictive validity and internal reliability of DriveSafe.

Method. A historical cohort study using retrospective descriptive analysis of DriveSafe scores and on-road driving performance for 898 drivers with cognitive and/or physical impairments referred for a driving assessment.

Results. Rasch analysis provided evidence for construct validity and internal reliability of DriveSafe. Goodness of fit statistics for all items were within the acceptable range. The test separated the participants into four groups with high participant and item reliability indices. Using a cut off score of 95/164, the sensitivity of the test was 81% and the specificity 90%. However, when coupled with clinicians' judgments of participants' insight, this improved to more clinically useful levels.

Conclusion. There is evidence for both reliability and construct and predictive validity of DriveSafe. The contribution and a priori measurement of insight requires further research.

Driving a motor vehicle is an integral part of daily life in modern western society enabling people to participate in work, leisure and family activities (Schultheis et al., 2002: 38; Sommer et al. 2004: 62). Driving, therefore, represents independence in mobility and other aspects of life resulting in improved quality of life (Barnes & Hoyle, 1995: 115; Korteling & Kaptein, 1996: 138). Older people who no longer drive report decreased satisfaction with life, loss of independence and personal identity and increased levels of depression (Stutts & Wilkins, 2003: 431). Basic competence in motor, sensory, perceptual and cognitive skills and an ability to integrate these skills in a rapidly changing environment are required to safely operate a motor vehicle (Coleman et al., 2002: 1415). Other factors such as experience and attitudes also contribute to this complex task (Lundqvist & Ronnberg, 2001: 171).

Medical conditions, disabilities, accidents and the aging process can cause changes in the requisite skills that potentially affect a person's ability to drive (van Zomeren et al., 1987: 698). As testing every driver's actual driving performance is both expensive and potentially dangerous, researchers have, for the past 20 years, examined a variety of clinical tests to identify a screening test that predicts on-road driving performance. Some researchers have used accident statistics and traffic violations as the measure of safe driving performance (e.g. Coleman et al., 2002; Katz et al., 1990). However, as accidents and violations represent a failure to drive safely and are relatively infrequent events, many researchers have preferred to use an assessment of actual driving performance. Despite some difficulties, assessments conducted by a trained driving assessor in real traffic, rather than on a closed circuit, are considered to most closely resemble everyday driving performance (Withaar et al., 2000: 488). Only those studies including an on-road assessment in real traffic will be considered here.

Many of the earliest studies did not find any relationship between neuropsychological tests and on-road driving performance (Brooke et al., 1992: 181; Galski et al., 1990: 711; Korteling & Kaptein, 1996: 144; van Zomeren et al., 1988: 94). Small sample sizes and differing diagnoses may have accounted for these results. Other studies found weak correlations between the two (Duchek et al., 1998: 1346; Fox et al., 1997: 951; Galski et al., 1992: 329; Schanke & Sundet, 2000: 119; Sivak et al., 1981: 482). Researchers concluded that neuropsychological tests could not be used as the sole basis for the determination of fitness to drive.

Several studies have yielded strong and significant relationships between specific off-road screening tests and on-road driving performance. The Stroke Driver Screening Assessment (Nouri & Lincoln, 1992: 280; Radford & Lincoln, 2004: 783), the Cognitive Behavioral Driver's Inventory (Engum et al., 1988: 43; Klavora et al., 2000: 704), the Motor Free Visual Perception Test, the Trailmaking Test B (Korner-Bitensky et al., 2000: 256; Mazer et al., 1998: 747) and the Useful Field of View test (Myers et al., 2000: 286) all yielded promising predictive results in an initial study. Recently, the Cognitive Behavioral Driver's Inventory (CBDI) was found to have sensitivity of 62% and specificity of 81% for failing the on-road assessment, which was considered to be insufficiently predictive of on-road performance to replace a full driving evaluation (Bouillon et al., 2006: 425). Further research using Useful Field of View (UFOV) test has focused on the association between poor performance on UFOV and involvement in at-fault crashes rather than on-road assessment results (Ball et al., 2006: 82).

In a recent study, a battery of 12 neuropsychological tests correctly predicted on-road failure for 100% of participants up to the age of 69 years with various diagnoses, but for participants 70 years and over, the battery was less sensitive and specific (McKenna et al., 2004: 334). Snellgrove developed and trialed the Maze Task on a group of 115 older drivers with mild cognitive impairment or early dementia and correctly predicted 78% of failures on-road (sensitivity) and 82% of passes (specificity) (Snellgrove, 2005: 28). These two studies represent the most accurate prediction of on-road performance. To date, neither of them has been replicated.

In the absence of a simple and predictive screening test most researchers recommend a comprehensive assessment of fitness to drive including both off-road screening tests and an on-road driving assessment. In most Western countries, specialist occupational therapists who conduct driving assessments use a comprehensive approach. This is the case in Australia where clients complete off-road screening tests of their visual and physical function, a pen and paper test of knowledge of road rules and a test to assess their awareness of the driving environment.

In assessing global awareness of the driving environment, DriveSafe is conceptually different to other tests currently being used and evaluated that test component visual processing or cognitive skills. It is widely used in Australia after being developed more than 12 years ago at the University of Sydney. Anecdotally this test provides useful information about a client's ability to drive and has good face validity. However, its psychometric properties had not been empirically analyzed. If this test accurately predicts those who are not fit to drive then the cost and potential risk of on-road assessment could be avoided for many clients. Thus, the purpose of this study was to examine the construct and predictive validity and internal reliability of DriveSafe.

Methods

Study Design

In this historical cohort study the medical files of all eligible participants were reviewed. Demographic information including age, gender and diagnosis, the results from DriveSafe and the on-road assessment and the clinicians conducting the assessment were recorded.

Participants

Data from all clients referred to two major driving rehabilitation centers in the Sydney metropolitan area, Calvary Rehabilitation and Geriatric Services (CRAGS) and Driver Rehabilitation and Fleet Safety Services, (DR&FSS) at The University of Sydney, over a 10 year period, were analyzed retrospectively (n = 898). The majority of the participants were assessed at University of Sydney's DR&FSS (58.6%). There were 641 men (71.4%) and 257 women (28.6%), with an age range of 16 to 93 years (mean age 52, SD 19.7). A wide spectrum of diagnoses was represented as illustrated in Table 1.

To be included in the study, participants had to have completed an off-road assessment, including DriveSafe, and an on-road assessment. Those clients who used an interpreter for the assessment or who had receptive or expressive aphasia were excluded from the study because they were unable to complete the standard measures.

Measures

DriveSafe consists of 15 images of the same rotary or roundabout (an alternative to a 4-way stop), projected on a screen to simulate the view through a windsheild, in which the number and position of pedestrians and vehicles vary (Refer to Figure 1). Participants are asked to observe each image for 3 seconds and when the image has been removed from the screen, to report details about the position and direction of travel of each pedestrian and vehicle in the slide. The images vary in complexity, requiring participants to report from 4 to 16 elements. The participants complete three practice images to ensure that they understand the instructions. Performance is recorded as a score out of 164. DriveSafe takes 20 minutes to administer and verbal responses are recorded by the clinician. If participants use hand gestures in providing answers they are requested to respond verbally but if unable to do so, the clinician accepts correct responses but notes that hand gestures were used. In the reported study, the test was administered according to the standard instructions using the standard scoring sheet. All clinicians were experienced driving assessors who had completed comprehensive training in administration of the test specifically focused on achieving high inter-rater reliability.

A 60-minute on-road driving assessment was completed by the same clinician within 1 week of the clinical assessment. The vehicle had automatic transmission, power steering and dual brakes. A registered driving instructor, in the passenger seat, gave directions and monitored safety while a registered driver-trained occupational therapist, sitting behind the instructor, recorded the participant's driving performance. The on-road assessment began in quiet suburban streets to allow the participant to become familiar with the vehicle controls before progressing to more demanding driving environments. Each center used a standard route unless vehicle modifications (such as hand controls) were required in which case the assessment continued in quiet suburban streets. The clinician recorded and categorized participant performance under the headings of observation, speed control, planning and judgment, vehicle positioning, vehicle control and reaction time. Any driving instructor interventions (e.g., use of the dual brake to prevent an accident) were recorded. The outcome of the assessment was determined as being either:

- Pass: safe and legal driving and no further intervention required
- Conditional Pass: safe and legal driving with restrictions on license (automatic vehicle only, limiting driving distance or time)
- Downgraded to a Learner's License: to undertake a series of driving lessons or
- Fail: failed to meet criteria for safe and legal driving and judged not to have the potential for improvement.

The criteria for failure were errors in all areas of driving or substantial errors in two or three areas and/or driving instructor intervention required to avoid a collision. For the purposes of statistical analysis, these categories were collapsed into a Pass (pass, conditional pass or downgrade to learner's license) or Fail category.

Data Analysis

Construct validity and internal reliability of DriveSafe were examined using Rasch modeling (Bond & Fox, 2001) with the WINSTEPS 3.58 program (Linacre, 2005). Rasch modeling constructs a linear measure from ordinal scores and assesses goodness of fit of both items and participants along a continuum, in this case, describing awareness of the driving environment. An item and participant map is generated in which the items are arranged in order of difficulty and participants are arranged in order of competence.

The program generates 2 pairs of fit statistics, infit and outfit statistics, expressed as mean square (MnSq) and standardized fit statistics that indicate how well the data from each item and participant conform to the assumptions of the Rasch model. The assumptions of the Rasch model are that easy items are easy for all people and that more competent people will perform better on all items. For adequate fit to the model, MnSq values of 1 ± 0.4 and standardized values of -2 to +2 were taken as acceptable for items and participants (Bond & Fox, 2001: 34). Fit statistics below and above this level indicate too little or too much variation respectively. Items with fit statistics in excess of these acceptable limits should be considered for removal from the test (Garratt, 2003: 80). Ninety-five percent fit is desired and provides evidence of uni-dimensionality. Characteristics of participants who did not fit the model were investigated to determine if there were any patterns that might explain why the test might be working differently for these individuals.

A principal component analysis was also conducted. When used with Rasch modeling, the purpose of principal component analysis is to examine uni-dimensionality of the scale rather than to construct variables as in standard factor analysis. If the empirical variance closely matches the modeled variance then the test fits the expectations of the Rasch model; if the unexplained variance from the first factor is less than 3 Eigenvalue units then this provides additional evidence that the test is uni-dimensional (Linacre, 2005).

Rasch modeling aslo produces reliability estimates for both participants and items. A separation statistic provides evidence of internal reliability or the ability of the instrument to separate groups of participants into levels of ability. In order to conclude that differences in the measure are due to real differences in the extent to which participants possess the trait (awareness of the driving environment) and not to error of measurement, the separation statistic should be 2.00 or greater. The participant reliability index (Cronbach's alpha equivalent) and the item reliability index, or the replicability of placement of participants or items respectively along the continuum should be .80 or better. Additionally, point measure correlation coefficients should be positive and large enough to show a strong relationship between the item and the construct (>.50).

To ensure that the test items function similarly (Badia et al., 2002) for participants with different relevant characteristics (e.g., gender) a differential item function analysis (DIF) is computed by Winsteps. Items that are functioning significantly differently ($p \leq 0.05$) for participants with specific characteristics are identified and could be considered for removal from the test. When more than one rater is scoring the test it is possible to use DIF analysis of raters to determine if there are any unexpectedly large differences in the way in which raters score each item on the test. In the absence of traditional inter-rater reliability data, this analysis provides evidence for inter-rater reliability.

Finally, the predictive validity of DriveSafe was calculated. Data from all participants were randomly allocated to two groups. A discrimination score was determined for one group using descriptive statistics (i.e. specificity, sensitivity, positive and negative predictive values). The accuracy of the score was then tested in the second group. The most useful test will have a positive predictive value approaching 100%, minimizing the number of drivers who failed the clinical test but passed the on-road test (false positives).

Results

Construct Validity

All items had fit statistics within the acceptable range suggesting that the items conformed to the assumptions of the Rasch model. The map of items and drivers (Figure 2) demonstrated that the spread of participants was greater than the spread of items, indicating that the test does not assess the full range of driving competence. Specifically, the most competent participants were insufficiently assessed. However, those participants below the mean or who were least competent were adequately assessed and this is the group of greatest concern. There were also some items that did not contribute to the ability of the test to differentiate between participants with varying levels of skill. Four levels contained multiple items and given that all items contribute in similar ways to the overall construct, one or more items in each level could be considered for removal from the test to reduce administration time.

Principal components analysis demonstrated that the empirical variance (86.2%) closely matched the modeled variance (82.3%). Furthermore, the unexplained variance by the first factor was only 1.6 Eigenvalue units. These provide further evidence that the test is uni-dimensional.

Internal Reliability

The test separated the participants into four groups (a model separation of 4.58) with a participant reliability index (Cronbach's alpha equivalent) of 0.94. All items had point measure correlation co-efficients between 0.63 and 0.87. The item reliability index was 1.00. Data from approximately 91% of participants conformed to the expectations of the Rasch model. Those whose data failed to fit the model were examined further to determine if there were any patterns that would explain the results. Five percent of participants had unacceptably high fit statistics indicating erratic responses and 4% had unacceptably low fit statistics indicating overly predictable responses (i.e., too Guttman-like) or more probably that their scores were within a narrow range. Although participants with low fit statistics are of less concern that those with unacceptably high fit statistics, all participants with mis-fitting responses shared some similar characteristics. They were younger (>60% were < 65 years), demonstrated good awareness of the driving environment (>70% scored 110/164 on DriveSafe) and drove safely (>80% passed the on-road assessment). This analysis indicates that there is no need for further more sophisticated modifications to the measure (Smith, 1996).

The DIF analysis revealed that there was no significant difference ($p \leq 0.05$) in the performance of participants on each test item with respect to gender. When the DIF analysis was applied to raters, it emerged that 4 of the 13 raters (E, I, J and L) had somewhat unusual scoring profiles on a few items (Figure 3). These raters were using DriveSafe in its earliest days. They assessed only a small number of the participants whose data did not conform to Rasch assumptions suggesting that overall they did score participants significantly differently ($p \leq 0.05$) than other raters. Thus the DIF analysis provides evidence for the inter-rater reliability of this instrument when used by trained raters according to standardized procedures.

Predictive Validity

The outcome of the on-road assessment was a “pass” for 394 participants (44%), a “conditional pass” for 171 participants (19%), a “downgraded license” for 196 participants (21.8%) and a “fail” for 137 participants (15.2%). When all participants who did not fail were included in the pass category, 761 participants (84.8%) passed.

A raw score cut off of ≤95/164 yielded a sensitivity of 82% (unsafe drivers who failed DriveSafe) and a specificity of 90% (safe drivers who passed DriveSafe), a positive predictive value (predicting those drivers who failed the on-road test) of 63% and a negative predictive value of 96% (predicting those drivers who passed the on-road test) in the first randomly allocated group (*n*=478). When this cut off score was tested on the second group (*n*=420) the sensitivity was 80%, specificity was 90% and the positive and negative predictive values were 55% and 97% respectively. Because this cut off score did not predict unsafe driving behavior as accurately as was expected a content analysis of the participants' files was conducted.

Lack of insight, or participants' decreased awareness of their driving ability, was a factor reported in most files, of participants who failed the on-road driving assessment. That is, clinicians documented that participants were unaware of driving errors when they were provided with feedback or that, in response to questioning, these participants reported that they had driven well despite driving instructor interventions. This was interpreted as being indicative of reduced insight. Thus we coded the clinician's judgment of the participant's insight as "intact" or "decreased" and added that to the predictive equation together with the test score. The descriptive statistics for several cut off scores were compared (refer to Table 3) and a cut off score of ≤110/164 together with decreased insight, yielded the best statistics: sensitivity of 83%, specificity of 99% and positive and negative predictive value of 97% in the first randomly allocated group. In the second group, the values were 87%, 100% and 98% respectively. Overall, only three participants (0.3%) predicted to fail the on-road driving test actually passed.

Discussion

The purpose of this study was to examine the construct and predictive validity and internal reliability of DriveSafe to determine how well test scores predict drivers who are unfit to drive. The findings yielded strong evidence for construct validity and internal reliability indicating that DriveSafe measures a theoretical construct related to driving and necessary for safe driving, namely, awareness of the driving environment. However, there are several redundant items that could be deleted to create a more efficient test without losing any discriminative power.

Used alone, DriveSafe predicted drivers who would pass an on-road assessment but was not as good at predicting those drivers who would fail the on-road assessment. That is, about 41% of drivers who failed DriveSafe (n=184) and were therefore predicted to fail on-road, actually passed. Because it would not contribute to drivers unfairly losing their licenses, this is a better outcome than the reverse (predicting a pass on-road and actually failing); however, it is higher than desirable. Nonetheless, when a judgment of participants' insight was used together with a discrimination score of 110/164 on DriveSafe, the predictive accuracy increased and 97.5% of participants who were predicted to fail the on-road assessment, failed the assessment.

The importance of insight for safe driving performance has been documented in the literature for healthy older drivers (Marottoli & Richardson, 1998: 332) and drivers with dementia (Cotrell & Wild, 1999: 155; Wild & Cotrell, 2003: 32) and traumatic brain injury (Brooks & Hawley, 2005: 173; Huchler et al., 2001: 286) although the precise nature of the relationship needs further study. From a theoretical perspective, insight underpins the widely accepted hierarchical model of driving performance (Michon, 1985). Drivers are unable to make strategic adjustments to their driving, such as avoidance of complex traffic if they are unaware of the need to do so (Ball et al., 1998: 320). Similarly, a recently documented, theoretically based, multi-factorial model suggests that safe driving behavior requires accurate self-monitoring of cognitive, sensory and physical function (Anstey et al., 2005: 60). A suggested flow chart that yields an accurate prediction of on-road performance is illustrated in Figure 4. This requires further research for investigation.

An important limitation of this study was that the clinicians reported the level of participants' insight based on their performance on both the off and on-road tests rather than on their performance in the off-road assessment alone. If insight is to be used with DriveSafe as a predictor of driving performance, then an accurate method of measuring the construct prior to the on-road assessment is required (Howorth & Saper, 2003: 121; Sherer et al., 2003: 60).

One of the strengths of DriveSafe is its clear face validity. Face validity refers to the suitability of an assessment in a practical situation and the confidence the users have in its efficacy (Anastasi, 1988). Many of the cognitive tests that have been used in previous research do not have face validity for testing driving, compromising the use of test results as a basis for cancellation of driving licenses. Nonetheless, asking participants who lack insight regarding their driving ability to accept a decision about license status without taking an on-road assessment may create new problems.

Limitations

As DriveSafe has not been previously psychometrically analyzed, there are no true inter-rater reliability statistics available for the test. It was not possible in this retrospective analysis to address this limitation using standard statistics. However, DIF analysis within Rasch modeling provided evidence for inter-rater reliability. Another limitation of DriveSafe is its reliance on language skills. If clients must rely on an interpreter then the test is likely not to be valid. Similarly, if clients have receptive or expressive language disorders then the test cannot be administered according to the standardized instructions. Future modifications of the test will need to be developed that eliminate the need for verbal responses. The influence of socio-economic or educational status on test performance has not been investigated.

The limitations common to retrospective studies, in general, were true of this study. There is an inherent selection bias in the sample as it represents only participants who were referred for driving assessment. There was no normative sample with which to compare the participants undergoing driving assessment. Additionally the on-road assessors were not "blind" to the participants' performance on DriveSafe because in the clinical setting the purpose of completing the off-road assessments is to inform assessors of deficits that may impact on on-road performance. The assessors could have been biased for or against certain participants. However, because this was a retrospective study, the assessors were unaware of the purpose of the study at the time they conducted the assessment (Bouillon et al., 2006: 426). Lack of information concerning reliability of the outcome measure (i.e., on-road driving performance) was a further limitation of the study. However, retrospective studies reflect clinical reality where it is not always possible to control for all the potential variables (Korner-Bitensky et al., 2000: 258). Furthermore the retrospective nature of this study enabled a large sample size, a characteristic rarely present in studies of driving performance.

Conclusion

Rasch analysis of DriveSafe provided evidence for the construct validity and internal reliability of the test. The test could be improved through the removal of redundant items. The test, with a clinician's judgment of insight, accurately predicted those drivers who are unsafe to drive. However, until a formal measure of insight is determined, use of the test as a sole basis for license cancellation is not supported. The predictive validity of the test prospectively and a suitable measure of insight require further research.

Acknowledgments

The authors thank the clinicians at DR&FSS and CRAGS for their support, Alexandra Wu and Virginia Deal for their assistance in data collection and entry and Dr Tim Luckett for his assistance with manuscript preparation.

Presented in part as a research poster at the 14th Congress of the World Federation of Occupational Therapists, July 2006, Sydney, Australia.

Suppliers

WINSTEPS, PO Box 811322 Chicago, IL 60681-1322

References

Anastasi, A. (1988). *Psychological Testing* (6th ed.). New York: Macmillan Publishing Company.

Anstey, K. J., Wood, J., Lord, S., & Walker, J. G. (2005). Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical Psychology Review, 25*, 45-65.

Badia, X., Prieto, L., & Linacre, J. M. (2002). *Differential item and test functioning*. Retrieved 26 September 2006, from http://rasch.org/rmt/rmt163g.htm

Ball, K., Owsley, C., Stalvey, B., Roenker, D. L., Sloane, M. E., & Graves, M. (1998). Driving avoidance and functional impairment in older drivers. *Accident Analysis & Prevention, 30*(3), 313-322.

Ball, K., Roenker, D., Wadley, V., Edwards, J., Roth, D., McGwin, G., et al. (2006). Can high-risk older drivers be identified through performance-based measures in a Department of Motor Vehicle setting? *Journal of American Geriatrics Society, 54*(1), 77-84.

Barnes, M. P., & Hoyle, E. A. (1995). Driving assessment - a case of need. *Clinical Rehabilitation, 9*, 115-120.

Bond, T. G., & Fox, M. C. (2001). *Applying the Rasch Model*: Mahwah, N.J.: L.Erlbaum.

Bouillon, L., Mazer, B., & Gelinas, I. (2006). Validity of the Cognitive Behaviour Driver's Inventory in predicting driving outcome. *The American Journal of Occuptional Therapy, 60*(4), 420-427.

Brooke, M. M., Questad, K. A., Patterson, D. R., & Valois, T. A. (1992). Driving evaluation after traumatic brain injury. *American Journal of Physical Medicine & Rehabilitation., 71*(3), 177-182.

Brooks, N., & Hawley, C. A. (2005). Return to driving after traumatic brain injury: a British perspective. *Brain Injury, 19*(3), 165-175.

Coleman, R. D., Rapport, L. J., Ergh, T. C., Hanks, R. A., Ricker, J. H., & Millis, S. R. (2002). Predictors of driving outcome after traumatic brain injury. *Archives of Physical Medicine & Rehabilitation., 83*(10), 1415-1422.

Cotrell, V., & Wild, K. (1999). Longitudinal study of self-imposed driving restrictions and deficit awareness in patients with Alzheimer Disease. *Alzheimer Disease and Associated Disorders, 13*(3), 151-156.

Duchek, J. M., Hunt, L., Ball, K., Buckles, V., & Morris, J. C. (1998). Attention and driving performance in Alzheimer's Disease. *Journal of Gerontology: Psychological Sciences, 53B*(2), 130-141.

Engum, E. S., Cron, L., Hulse, C. K., Pendergrass, T. M., & Lambert, W. (1988). Cognitive behavioural driver's inventory. *Cognitive Rehabilitation, September/October*, 34-46.

Fox, G. K., Bowden, S. C., Bashford, G. M., & Smith, D. S. (1997). Alzheimer's Disease and driving: Prediction and assessment of driving performance. *Journal of American Geriatric Society, 48*(8), 949-953.

Galski, T., Bruno, R. L., & Ehle, H. T. (1992). Driving after cerebral damage: a model with implications for evaluation. *American Journal of Occupational Therapy., 46*(4), 324-332.

Galski, T., Ehle, H. T., & Bruno, R. L. (1990). An assessment of measures to predict the outcome of driving evaluations in patients with cerebral damage. *The American Journal of Occupational Therapy, 44*(8), 709-713.

Garratt, A. M. (2003). Rasch analysis of the Roland Disability Questionnaire. *Spine, 28*(1), 79-84.

Howorth, P., & Saper, J. (2003). The dimensions of insight in people with dementia. *Aging and Mental Health, 7*(2), 113-122.

Huchler, S., Jorg, A., Golz, D., Kust, J., Netz, J., Homberg, V., et al. (2001). Evaluation of driving simulation. *Europa Medicophysica, 37*(4), 283-289.

Katz, R. T., Golden, R. S., Butter, J., Tepper, D., Rothke, S., Holmes, J., et al. (1990). Driving safety after brain damage: follow-up of twenty-two patients with matched controls. *Archives of Physical Medicine & Rehabilitation., 71*(2), 133-137.

Klavora, P., Heslegrave, R. J., & Young, M. (2000). Driving skills in elderly persons with stroke: comparison of two new assessment options. *Archives of Physical Medicine and Rehabilitation, 81*(6), 701-705,.

Korner-Bitensky, N. A., Mazer, B. L., Sofer, S., Gelina, I., Meyer, M. B., Morrison, C., et al. (2000). Visual testing for readiness to drive after stroke: a multicenter study. *American Journal of Physical Medicine and Rehabilitation, 79*(3), 253-259.

Korteling, J. E., & Kaptein, N. A. (1996). Neuropsychological driving fitness tests for brain-damaged subjects. *Archives of Physical Medicine & Rehabilitation., 77*, 138-146.

Linacre, J. M. (2005). *A user's guide to WINSTEPS Ministep Rasch-Model computer programs*. Retrieved 16 June 2005, from http://www.winsteps.com/winpass.htm

Lundqvist, A., & Ronnberg, J. (2001). Driving problems and adaptive driving behaviour after brain injury: A qualitative assessment. *Neuropsychological Rehabilitation, 11*(2), 171-185.

Marottoli, R. A., & Richardson, E. D. (1998). Confidence in, and self-rating of, driving ability among older drivers. *Accident Analysis & Prevention, 30*(3), 331-336.

Mazer, B. L., Korner-Bitensky, N., & Sofer, S. (1998). Predicting ability to drive after stroke. *Archives of Physical Medicine and Rehabilitation, 79*(7), 743-749.

McKenna, P., Jefferies, L., Dobson, A., & Frude, N. (2004). The use of a cognitive battery to predict who will fail an on-road driving test. *British Journal of Clinical Psychology, 43*, 325-336.

Michon, J. A. (1985). A critical view of driver behaviour models: what do we know, what should we do? In L. Evans & R. C. Schwing (Eds.), *Human Behaviour and Traffic Safety* (pp. 485-520). New York: Plenum Press.

Myers, R. S., Ball, K. L., Kalina, T. D., Roth, D. L., & Goode, K. (2000). Relation of Useful Field of View and other screening tests to on-road driving performance. *Perceptual and Motor Skills, 91*, 279-290.

Nouri, F. M., & Lincoln, N. B. (1992). Validation of a cognitive assessment: predicting driving performance after stroke. *Clinical Rehabilitation, 6*, 275-281.

Radford, K. A., & Lincoln, N. B. (2004). Concurrent validity of the Stroke Drivers Screening Assessment. *Archives of Physical Medicine and Rehabilitation, 85*(2), 324-328.

Schanke, A. K., & Sundet, K. (2000). Comprehensive driving assessment: neuropsychological testing and on-road evaluation of brain injured patients. *Scandinavian Journal of Occupational Therapy, June*, 113-121.

Schultheis, M. T., Matheis, R. J., Nead, R., & DeLuca, J. (2002). Driving behaviors following brain injury: self-report and motor vehicle records. *Journal of Head Trauma Rehabilitation., 17*(1), 38-47.

Sherer, M., Hart, T., & Nick, T. G. (2003). Measurement of impaired self-awareness after traumatic brain injury: a comparison of the patient competency rating scale and the awareness questionnaire. *Brain Injury, 17*(1), 25-37.

Sivak, M., Kewman, D. G., & Henson, D. L. (1981). Driving and perceptual / cognitive skills: Behavioural consequences of brain damage. *Archives of Physical Medicine & Rehabilitation., 62*, 476-483.

Smith, R. M. (1996). *Polytomous Mean-Square fit statistics*. Retrieved 6 November 2006, from http://www.rasch.org/rmt/rmt103ahtm

Snellgrove, C. A. (2005). *Cognitive screening for the safe driving competence of older people with mild cognitive impairment or early dementia.* Canberra: Australian Transport Safety Bureau.

Sommer, S. M., Falkmer, T., Bekiaris, E., & Panou, M. (2004). Toward a client-centred approach to fitness-to-drive assessment of elderly drivers. *Scandinavian Journal of Occupational Therapy, 11*(2), 62-69.

Stutts, J. C., & Wilkins, J. W. (2003). On-road driving evaluations: A potential tool for helping older adults drive safely longer. *Journal of Safety Research, 34*(4), 431-439.

van Zomeren, A. H., Brouwer, W. H., & Minderhoud, J. M. (1987). Acquired brain damage and driving: a review. [Review] [35 refs]. *Archives of Physical Medicine & Rehabilitation., 68*(10), 697-705.

van Zomeren, A. H., Brouwer, W. H., Rothengatter, J. A., & Snoek, J. W. (1988). Fitness to drive a car after recovery from severe head injury. *Archives of Physical Medicine & Rehabilitation., 69*, 90-96.

Wild, K., & Cotrell, V. (2003). Identifying driving impairment in Alzheimer Disease: A comparison of self and observer reports versus driving evaluation. *Alzheimer Disease and Associated Disorders, 17*(1), 27-34.

Withaar, F. K., Brouwer, W. H., & van Zomeren, A. H. (2000). Fitness to drive in older drivers with cognitive impairment. *Journal of the International Neuropsychological Society, 6*, 480-490.

Table 1: Participant Diagnoses (*n*=898)

Diagnosis	**Number**	**Percentage**
Orthopedic/Spinal Injury	366	40.76
Neurological/Cognitive Impairment (including CVA, dementia)	255	28.4
Multiple Diagnoses (including neurological/cognitive impairment)	148	16.48
Traumatic Brain Injury	60	6.68
Miscellaneous Diagnoses	46	5.12
Vision Impairment	23	2.56

Table 2: Item Statistics, Items arranged from Hardest to Easiest

Item	Details to be recalled	Measure Score	Infit MnSq	Outfit MnSq	Point Meas Correlation
2	12	58.3	1.31	1.37	.81
9	16	54.6	1.02	1.06	.85
3	8	53.3	.94	.97	.77
5	12	53.3	.83	.84	.84
6	16	53.0	.84	.86	.86
10	16	52.9	.97	.98	.86
13	8	52.4	.92	.94	.78
8	16	51.2	.92	.96	.85
1	12	49.7	1.05	1.16	.75
11	12	46.9	.94	.93	.79
7	8	46.0	1.04	1.00	.71
14	8	45.8	.82	.91	.73
12	8	45.3	.94	.91	.74
15	4	44.6	.83	.82	.63
4	8	42.6	1.29	1.05	.68

Table 3: Descriptive Statistics for Cut off Scores on DriveSafe with Decreased Insight

Cut off Score (Raw score)	**100**	**105**	**110**	**115**
Cut off Score (Measure score)	**56.1**	**57.5**	**58.9**	**60.3**
Group 1 (*n*=478)				
Sensitivity	82	83	83	84
Specificity	100	99	99	99
Positive Predictive Value	100	97	97	97
Negative Predictive Value	96	97	97	97
Group 2 (*n*=420)				
Sensitivity	76	82	87	89
Specificity	100	100	100	99
Positive Predictive Value	98	98	98	96
Negative Predictive Value	97	97	98	98
Total (*n*=898)				
Sensitivity	79	82.5	85	86.5
Specificity	100	99.5	99.5	99
Positive Predictive Value	99	97.5	97.5	96.5
Negative Predictive Value	96.5	97	97.5	97.5

Note: The values represent percentages.

Figure 1: Example of DriveSafe item

Figure 2: Map of Items and Drivers

```
  Raw     Measure                 Drivers    Items
  Score    Score             more capable  |  hardest
  >162      100                        .    +
                                            |
                                            |
                                            |
                                       .    |
                                            |
   160       90                        .    +
                                       .    |
                                       .    |
                                       .  T |
                                      .#    |
                                     .##    |
   154       80                       .#    +
                                    .###    |
                                  .#####    |
                                    ####    |
                                  .#####  S |
                                 .######    |
   141       70              .#########     +
                            .##########     |
                          .############     |
                             .#########     |
                                 .###### M  |
                               .########    |
   114       60                .########    +
   110                        .#########    | T   2                  Cut off –  Decreased insight
------------------------------------------------------------------------------------------------
                                 .######    |
                                .#######    | S   9
    95                             .#### S  |    10     3     5     6      Cut off – Test only
------------------------------------------------------------------------------------------------
                                  .#####    |    13     8
    78       50                    .####    + M   1
                                    .###    |
                                     .##    |    11     7
                                      ##    | S  12    14    15
                                      ##    |     4
                                       # T  | T
    47       40                       .#    +
                                       .    |
                                       .    |
                                       .    |
                                       .    |
                                       .    |
   <24       30                        .    +
                            Less capable    |  easiest

Each '#' is 6,   M=mean,   S=one standard deviation,   T=two standard deviations
```

Figure 3: Rater Differential Item Functioning (DIF) Analysis

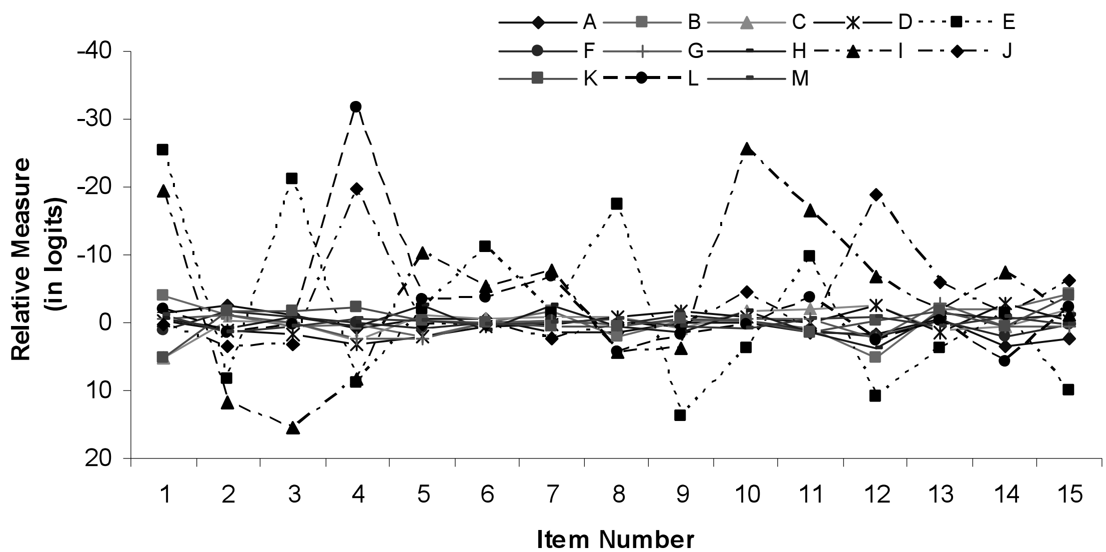

Figure 4: Flowchart for Predicting Driving Outcome