J Rehabil Med 2006; 38: 273–279

P-DRIVE: ASSESSMENT OF DRIVING PERFORMANCE AFTER STROKE

Ann-Helen Patomella, OT, MSc, Kerstin Tham, OT, PhD and Anders Kottorp, OT, PhD

From the Neurotec Department, Division of Occupational Therapy, Karolinska Institutet, Stockholm, Sweden

Objective: **To investigate aspects of validity and stability of Performance Analysis of Driving Ability (P-Drive), for people with stroke when used in a driving simulator.**
Design: **A cross-sectional observational study.**
Subjects: **The study included a consecutive series of 101 participants with stroke referred for evaluation or selected from a stroke registry.**
Methods: **P-Drive was used to observe driving performance in order to assess driving ability. P-Drive comprises 20 items assessing the quality of the participant's driving performance. Aspects of validity and reliability in P-Drive were evaluated using Rasch statistics.**
Results: **The items (95%) and participants (97%) demonstrated acceptable goodness-of-fit and met statistical expectations according to the Rasch model. The results support internal scale validity and person response validity. P-Drive could separate the participants with different driving abilities and the standard errors were within reasonable criteria for drivers with a moderate-to-low ability to drive.**
Conclusion: **The findings from this study indicated that P-Drive is an assessment tool with properties of internal scale validity, person response validity, and which also contains aspects of reliability in relation to precision of the estimates and separation. P-Drive seems to be a valid and stable assessment tool for assessing the driving ability in a simulator of people with stroke.**

Key words: driving ability, simulator, Rasch analysis, occupational therapy, performance evaluation.

J Rehabil Med 2006; 38: 273–279

Correspondence address: Ann-Helen Patomella, Karolinska Institutet, Department Neurotec, Division of Occupational Therapy, Alfred Nobels Allé 23, SE-141 83 Huddinge, Sweden. E-mail: ann-helen.patomella@ki.se

Submitted August 8, 2005; accepted February 13, 2006

INTRODUCTION

People recovering from stroke are often interested in returning to driving, which is an understandable desire since many people need to drive in order to manage everyday life. On the other hand, the ability to drive may be reduced by stroke (1, 2). Physicians in rehabilitation medicine, together with allied rehabilitation professionals, commonly occupational therapists, are often called upon to determine the patient's ability to drive as a part of preparation for their discharge to the community (1, 3). On-road evaluation is often seen as the optimal method for assessing driving ability, but may be neither sufficiently challenging nor safe enough to detect incompetent driving ability. Demanding driving actions are best studied in an environment that is challenging yet safe for both the driver and examiner, as in the controlled environment of a simulator (4).

Nevertheless, driving evaluations in simulators have been criticized for being difficult to interpret, due to the absence of valid outcome scores. Simulators have also been criticized for lack of interactive technology (1, 5). The advantage of using simulators is the ability to use challenging situations that will capture incompetent driving ability, which could have been unsafe and also unlikely to happen on the day of an on-road evaluation (4). In this study, the combination of a driving simulator with an advanced interactive technique that is close to real driving was used to observe driving performance. So far, there is no specific assessment tool for assessing driving ability as performed in a simulator. Criticism could be made of on-road assessments used in research due to the absence of a theoretical frame of reference and because such assessments may consist of multiple constructs, sometimes even summed into raw scores (6–8). Ordinal scales and raw scores may be limited with respect to different aspects of validity and there is a need for ratio or interval scales to improve measurement, inference and prediction in rehabilitation medicine (9) as well as for evaluation of driving ability.

Performance Analysis of Driving Ability (P-Drive) is an assessment tool that generates equal-interval measures of driving ability and was developed recently for assessing driving ability in the context of a technically advanced and interactive simulator (10). The development of P-Drive was based on a theoretical frame of reference with 3 theoretical domains. First, measuring performance in actions rather than underlying capacities was considered as the primary focus using a top-down approach (11, 12) for the activity of driving. Secondly, the concept of driving ability, as measured in P-Drive, was based on actions that reflected the performance of driving and the items were developed using an activity analysis (13). Lastly, the hypothesized hierarchical order of the items was based on the model of Michon (14) and on theories of attention and information processing related to driving (2). The assumption was that items demanding attention and fast information processing were expected to be more difficult than items concerning general operation of the car (2). Also, items needing tactical decision-making were expected to be more difficult than decisions on operating the car (14). The aim in developing P-Drive was to create a useful tool for clinical rehabilitation when assessing people in a simulator.

DOI: 10.1080/16501970600632594

The results from a pilot study ($n = 31$) of P-Drive using a Rasch analysis, revealed evidence of validity in terms of both item and person goodness-of-fit statistics, indicating unidimensionality and person response validity (10). The hierarchy of the items was also found to be logical and in-line with assumptions made from previous research in the field of driving (2). The results, however, revealed a high mean standard error (SE) for the participants, probably due to the fact that the sample was skewed by a majority of drivers with a high level of driving ability. This pilot study indicated a need for further studies, since the sample was small and homogeneous, motivating the present study with a sample that is larger and will presumably include people with a wider range of impairments. The encouraging results of the pilot study also implied a need to investigate the stability of the assessment tool in a sample of people with stroke. The overall purpose of this study was, therefore, to determine the stability of different aspects of validity of P-Drive when used to assess driving ability in people with stroke. In this study the following research questions were addressed. (*i*) Do the participants demonstrate acceptable goodness-of-fit when assessed with P-Drive, indicating *person response validity*? (*ii*) Do the items in P-Drive demonstrate acceptable goodness-of-fit, indicating *internal scale validity*? (*iii*) Does P-Drive *separate* between the abilities of the participants? (*iv*) Are the ability measures of the participants associated with reasonable SE? (*v*) Do the item difficulty calibrations in P-Drive demonstrate acceptable stability across specific subgroups?

METHODS

Participants

The present study was approved by the Ethical Committee of Northern Stockholm. The study included 101 participants with stroke. The participants were recruited for the study from the referrals for a driving evaluation made by physicians in the Stockholm area to the Unit of Traffic Neurology, Karolinska Hospital during 2002–03 ($n = 74$). In addition, all the people with stroke on the Stockholm Stroke Registry during the period who had ticked a box indicating that they were active drivers before the stroke were invited to participate. Inclusion criteria for the study were: (*i*) a diagnosis of stroke, (*ii*) not getting nauseous during simulator driving, (*iii*) holding a driving license and have been an active driver before the stroke, and (*iv*) have given informal consent for participation in the study. A consecutive series of participants who met the criteria were included in the study. The characteristics of the participants are described in Table I. The sample size was set to at least 100 participants in order to obtain stable item calibrations (15).

Table I. *Clinical characteristics of the 101 participants*

Characteristics	Mean (SD)
Persons (*n*)	
Men	88
Women	13
Age (years)	61.9 (10.1)
Time since onset (months)	13.4 (13.9)
Range, median	2–60, 10

SD = standard deviation.

Fig. 1. Argus driving simulator. Photograph: Ann-Helen Patomella.

Study context

The interactive, realistic and technically-advanced simulator used in this study was situated at the Karolinska University Hospital (Stockholm, Sweden). The simulator was developed and used by clinical occupational therapists to observe driving performance after stroke. The simulator consists of a real, but truncated car (see Fig. 1). As in real driving, the driver has to operate the pedals, turn the steering-wheel, etc. When required the car can be adapted with technical devices such as a spinner-knob and automatic gears. The driving program is projected onto 3 large screens with a field-of-view of 135°. An audio system generates synchronized sounds, such as the engine running and radio messages. The test-program consisted of 70 traffic situations, each presenting a different challenge. The drive took about 40–60 minutes, depending on the speed and the number of mistakes made. The simulator required the driver to deal with the type of situations that one would frequently encounter when driving a car, and had been designed specifically to involve situations assumed to be particularly challenging for people with stroke.

Assessment tool

P-Drive was used to score the participants' driving performance as observed when driving in the simulator. P-Drive had been slightly restructured since the first study (10). One item ("following radio message") was redundant and could be removed. Since the scale showed promising results of scale and person validity, it was used in the same manner as in the pilot study. Whilst driving in the simulator, each traffic situation that occurred during the test required the driver to perform a number of actions. There were 20 items defined in P-Drive, sorted into 4 subgroups: manoeuvring, orientating, obeying and responding to traffic rules and paying attention. The items were scored using a detailed and structured manual and a 4-point criterion referenced rating scale based on the quality of driving performance, where 4 = competent, 3 = questionable, 2 = problem and 1 = incompetent. During a complete driving test using the simulator's test-program, all of the items in P-Drive were observable several times and in different challenging situations (for examples of the scoring, see Table II).

Procedure

The evaluation started with a short review of the participant's medical and driving history. All of the participants were given the chance to practice in the simulator using a special practice-program that allowed the participant to get used to the car, that is, to the pedals, gear box and steering. The participant decided when no more practice was needed (commonly after 15–30 minutes). Thereafter, the participants underwent a driving evaluation using the test-program in the simulator. An occupational therapist (A-HP) with previous experience of performing driving evaluations scored each participant's driving performance using P-Drive. The scoring was structured using the criteria from the P-Drive manual and took place directly after the test drive.

Table II. *Examples of traffic situation and items scored*

Traffic situation	Items (actions) scored in P-Drive
Participant collided with a car coming from the right (had no right-of-way)	Paying attention to the right (score 1) Giving right-of-way (score 1) Controlling speed (scored depending on the way the participant adapted the speed for the oncoming crossing)
Participant turns competently into a slipway onto the highway, directed by an earlier radio message	Following instruction (score 4) Paying attention to signs providing information (score 4) Finding the way (score 4) Using indicator (score 4) Steering (score 4)

P-Drive = Performance Analysis of Driving Ability.

Data analysis

Computer application of Rasch analysis, Winsteps (16) was used to analyse the P-Drive data for the participants. The Rasch rating scale model used in the study could be stated as:

Log $[P_{nik}/P_{nik-1}] = B_n - D_i - F_k$, were
P_{nik} = probability of a person n being assigned score k on action item i.
P_{nik-1} = probability of a person n being assigned score $(k-1)$ on action item i.
B_n = ability measure of person n.
D_i = calibration of the difficulty of action item i.
F_k = the difficulty of rating scale step k relative to step $(k-1)$.

The Winsteps program generates goodness-of-fit statistics, which can be used to evaluate the extent to which the data conforms to the model stated above. When the goodness-of-fit statistics for the participants are acceptable, they provide evidence of *person response validity* for the participants. When the goodness-of-fit statistics for the items are acceptable, they provide evidence of *unidimensionality* or *internal scale validity* (17). For this study, both the mean square residual (MnSq) and the associated standardized z-value were used to evaluate goodness-of-fit. Infit and outfit mean square residuals between 0.6 and 1.4, associated with z-values larger than -2 and less than 2, were set as criteria for an acceptable goodness-of-fit (18). In accordance with other studies (19, 20), it is generally expected that no more than 5% of the participants or the items will fail to fit the model by chance. The effectiveness of the rating scale was determined by investigating the distribution of the observations for each category in the scale, noting whether disordered categories (thresholds) occurred and by the randomness of the choice of categories (21). The criteria used when determining randomness were outfit mean square measures of less than 2.0 (21). In addition, a principal components analysis of the items was performed to investigate the local independence of P-Drive and to determine whether any additional dimension was present in the dataset (22). The criterion for unidimensionality suggests that the first component should explain at least 20% of the variance (23) and the second one should not explain more than 5% of the variance.

The person separation reliability was investigated to determine if P-Drive was able to separate the participants according to their different levels of (driving) ability. For the scale to be able to distinguish between 2 or more groups in the sample, the person separation reliability should be at least 0.8 (24). In addition, the SE describes the statistical reproducibility of the measures, and a guiding principle for the SE in the person measures had previously been set at ≤ 0.30 logits (20). In this study, a SE analysis was performed in subgroups based on the level of driving ability, according to a recent pilot study investigating P-Drive as an outcome tool (25).

An assumption in instrument development using Rasch is that item calibration remains statistically stable across different groups (26, 27). In order to investigate the stability of the item hierarchy, an analysis of differential item functioning (DIF) was conducted for different groups within the sample and plotted with 95% confidence intervals (CI) (28). An item outside the CI would indicate that this item was relatively harder or easier to perform when used with a different sample of people. The criterion for calculating DIF was that the sample size for each group had to be at least $n = 30$ to generate stable item calibrations (15). The subgroups chosen for this study were: (*i*) persons who had had a right hemispheric cerebral vascular accident (RCVA) ($n = 24$) and persons who had had left cerebral vascular accident (LCVA) ($n = 39$), (*ii*) persons of 60 years of age or younger ($n = 54$) and persons older than 60 years ($n = 47$), (*iii*) persons with a better driving ability (≥ 1.4 logits on P-Drive) ($n = 51$) and persons with a worse ability to drive (< 1.4 logits on P-Drive) ($n = 50$).

RESULTS

Person response validity

Of the 101 participants, 98 demonstrated acceptable goodness-of-fit (97%) to the Rasch model, indicating acceptable person response validity. Figure 2 shows the targeting of the persons and items. The different response thresholds of the items are visible and show that most participants scored high on the rating scale.

Internal scale validity

Nineteen of the 20 items demonstrated acceptable goodness-of-fit (95%), and item 4, "controlling speed", (Infit MnSq 0.54 and z -4.2) did not fit the model expectations. We concluded, therefore, that the scale could be viewed as unidimensional, supporting acceptable internal scale validity. With the intention of analysing the impact of item 4 on the person ability measures another Rasch analysis of the 19 items that demonstrated acceptable goodness-of-fit was performed, revealing no difference in the person's ability measures.

When examining the categories of the rating scale, we found that there were irregularities in the use of the categories: a score of 4 had been used 56% of the times and a score of 1 only 7% (Table III). However, no step disordering was found in the scale category measures, indicating that the scale categories worked as intended. The outfit mean square was less that 2.0, indicating randomness in the choice of categories (21).

The principal components analysis revealed that the first component accounted for 64% of the variance indicating unidimensionality, and the second component explained about 6% of the variance, which is slightly higher than expected.

Separation of the participants

The person separation reliability of the participants was 0.84 and person separation index was 2.30, indicating that P-Drive was able to separate the participants' driving ability into different strata.

SE in the estimates

The mean SE for the participants was larger (mean = 0.42) than the criteria set. The result was in line with the result from the pilot study where the mean SE was 0.39. Further investigation of the SE revealed that participants with a good driving ability had higher SE than drivers in the middle and the lower end of

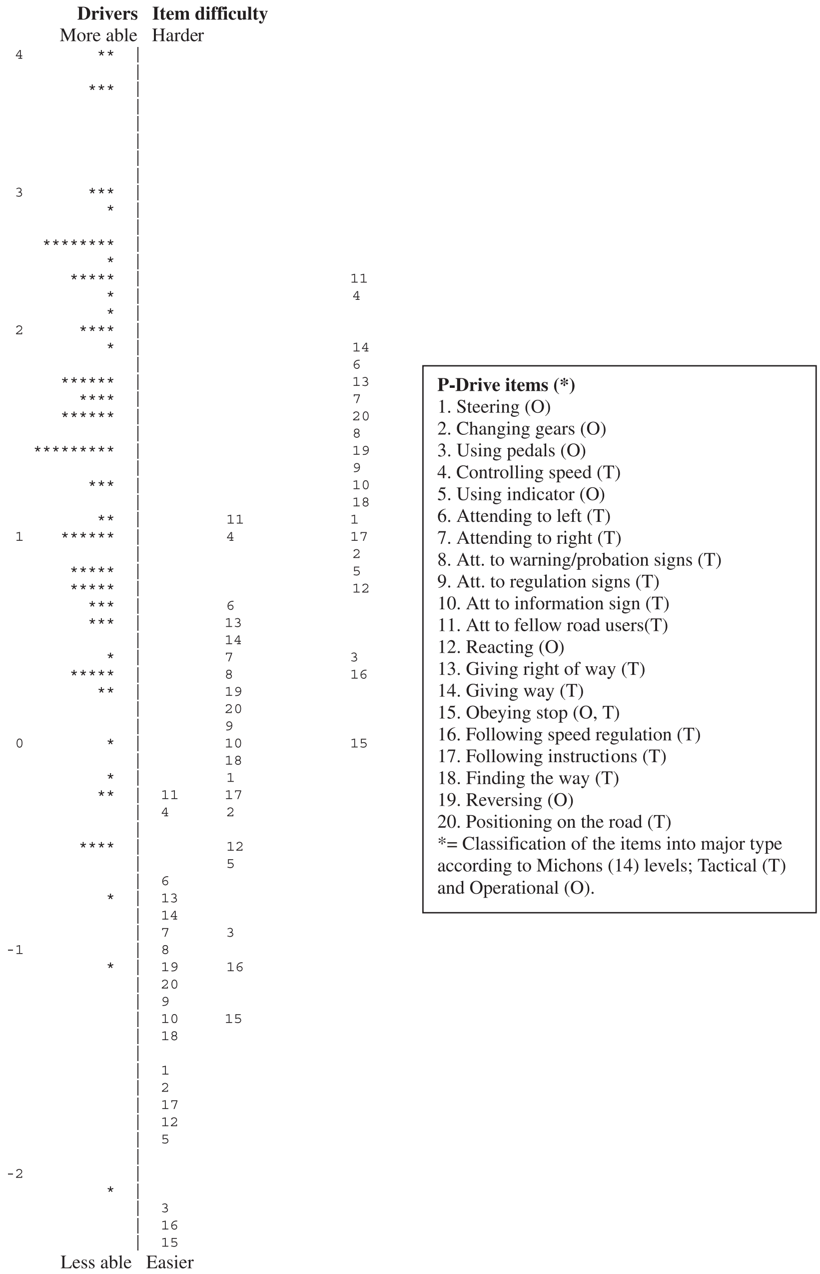

Fig. 2. Targeting of Performance Analysis of Driving Ability (P-Drive): person ability measures in relation to item difficulty calibrations including thresholds for response categories on the rating scale (1–2; 2–3; 3–4). Higher measure indicates higher person ability and more difficult item.

the scale (Table IV). The less able the driver, the lower the SE, suggesting that P-Drive is able to more precisely estimate the abilities of moderate-to-poor drivers than more able drivers.

Differential item functioning

The samples with right and left stroke (CVA) revealed no detectable difference in item calibration (Fig. 3). As only minor differences could be detected between older and younger drivers respectively drivers with high and low ability to drive, the conclusion remained that the overall item hierarchy was stable across the subgroups.

DISCUSSION

In summary, P-Drive seems to be a valid and stable assessment tool for assessing driving ability in people with stroke with the use of a simulator. The finding from this study indicated that P-Drive is an assessment tool with properties of internal scale validity, person response validity and also sustained aspects of reliability in relation to precision of the estimates. Furthermore, P-Drive seems to be able to differentiate between people with different driving abilities.

Table III. *Rating scale measures*

Scale category	Frequency (%)	Average category measure	Outfit MnSq
1. Incompetent	146 (7)	−0.56	0.85
2. Problem	231 (12)	0.35	0.94
3. Questionable	452 (23)	1.09	1.18
4. Competent	1100 (56)	1.89	1.03

MnSq = mean square residual.

Table IV. *Mean standard error (SE) in different strata of the sample*

Persons (*n*)	Driving ability measure	Mean SE
36	Over 1.7 logits	0.63
15	1.7–1.4 logits	0.34
50	Lower than 1.4 logits	0.28
		Total mean 0.42

Internal scale validity was found in P-Drive, however, one item failed to fit the Rasch model. Item 4, "controlling speed", did not meet the goodness-of-fit criterion with a mean square of 0.54, indicating less variation in the observed response pattern than was modelled (27). Further examination of the raw scores of the responses for this item revealed that it was too predictable. A raw score of 2 were given to almost all participants despite the fact that they had different types of difficulties in this item (i.e. both those with difficulty to control speed upward and downwards were given a score of 2). Assessment of the complexity of driving ability requires the ability to detect and separate different performance difficulties. The difficulty in controlling speed upward and downward may cause different driving performance difficulties and may be due to various kinds of dysfunction in the underlying capacities of the participant. For example, impulsive driving behaviour may result in incompetent and unsafe driving due to high speed (29), while reducing speed is an adaptive behaviour that may be due to self-awareness and motivation to drive safely (30). Also, in the clinical setting there have been examples of drivers who have not been able to follow the pace of the traffic and the exaggerated adaptation of reduced speed has resulted in an incompetent driving performance instead of safety. These variations of driving performance should be captured in P-Drive and item 4, "controlling speed", should preferably be modified instead of removed. One way to gain greater variation in the assessment of the ability to control speed could be to divide the item into 2 new items; "reducing speed" and "keeping speed" and then investigate the goodness-of-fit.

The person response validity was satisfactory and the items in P-Drive could separate the people in the sample into different

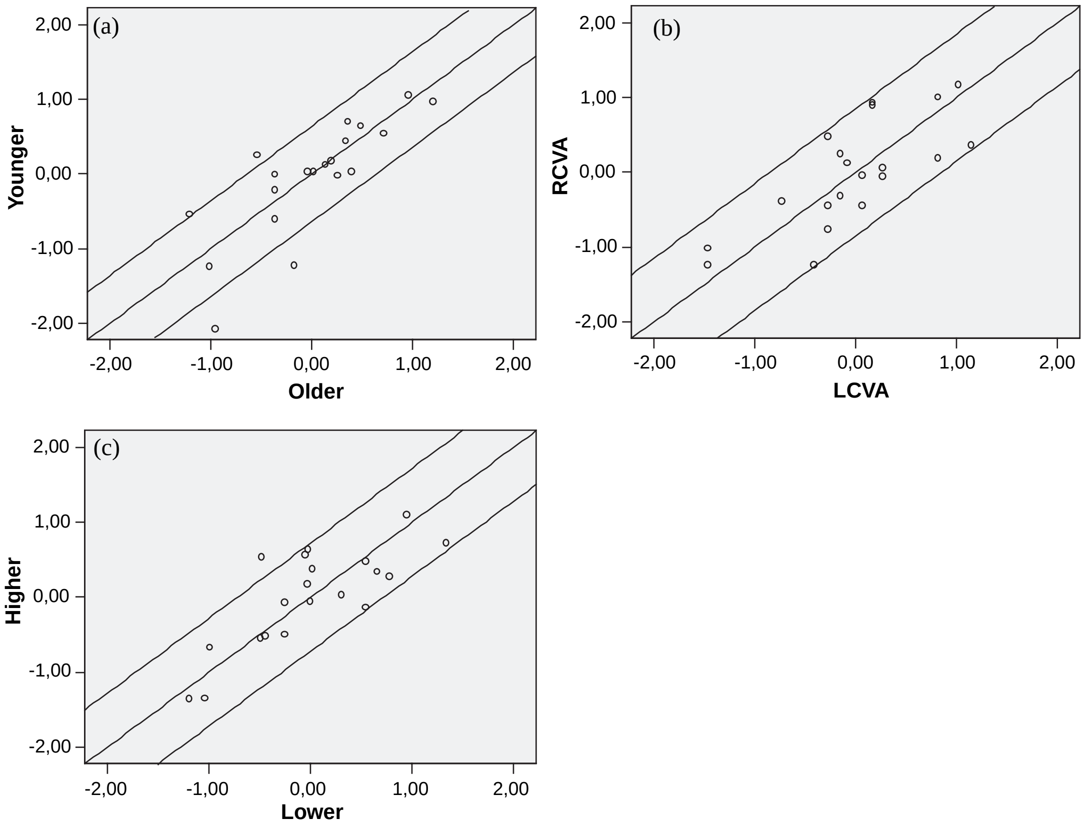

Fig. 3. Differential item functioning of: (a) younger (≤60 years) and older (>60 years) drivers; (b) drivers with right and left hemisphere stroke (RCVA and LCVA, respectively); and (c) drivers with higher (≥1.4 logits) and lower (<1.4 logits) ability to drive. The area between the two outer lines indicates the 95% confidence interval.

levels of ability. However, 3 participants did not fit the Rasch models expectations. Further examination of these participants' responses revealed that they all had low scores on specific items, but otherwise performed well. As an example, one of these participants had problems only with the items concerning situations needing attention to the left side of the traffic environment. This was a participant that had a right hemisphere stroke and initial sequelae of neglect. The unexpected response pattern could therefore be related to remaining symptoms from the neglect and therefore not viewed as a threat to person response validity.

High mean SE of the participants was reported in this study. The high SE for the drivers with a high ability indicated that we could expect noticeable different estimates of these participants if we reassessed their driving ability. Fortunately, the SE for the participants with moderate and low ability to drive were within reasonable limits (20). The SE for these participants is of greater importance since they are in the zone where they are at risk of failing the driving evaluation and there is a need for a high precision in the estimates generated. Similar problems with high SE have been reported in a study investigating on-road driving (8).

The results from this study indicated that the item hierarchy of P-Drive (see Fig. 2) was in line with the theoretical frame of reference used (2, 14). In other words, items requiring great attention and rapid information processing were the most challenging, such as item 11, "paying attention to fellow road users", and those concerning manoeuvring were the least challenging. Also items needing tactical decisions were more challenging than those concerning only operational decisions.

A methodological limitation of the study was the over-representation of male participants (87%), yet the inclusion criteria should not necessarily favour male inclusion. In Sweden there are no gender differences in stroke (31) or in licensing (46% of driving licence holders are women (32)) and could not be the reason for the over-representation. Similar over-representations are found in other studies (Canada and The Netherlands) measuring driving ability, with males representing 75–79% of the study population (6, 8). The reason behind the male dominance in referrals is probably complex and would be interesting to analyse further. One possible reason could be that men are more strongly motivated to return to driving. Thus, it is important that referring physicians are aware of the under-representation of women in referrals for driving evaluations. It would have been interesting to investigate DIF between men and women, but in order to obtain stable item calibrations there is a need for more women in the sample (16).

However, a pass or fail in a driving evaluation should not be based only on P-Drive, but rather on a consensus of the patient's medical condition, neuropsychological status, experience, judgement, etc. There is a need to investigate and develop cut-off criteria for P-Drive. In a recent pilot study preliminary cut-off criteria were developed using an on-road evaluation as criterion (25). This study was very small ($n = 27$), and for developing definitive cut-off a larger sample is needed. The use of on-road driving evaluation as a criterion for standardizing assessment tools for driving is common. However, a normal on-road driving test is seen as a rough estimate of a participant's ability and it is unusual that the driving is challenging enough to detect risky behaviour due to cognitive impairment (33). In the future the criterion for setting cut-off for driving outcome should be based on the total team-decision rather than a normal on-road evaluation.

P-Drive has so far been developed and investigated for its validity when measuring driving ability in a simulator. Unfortunately, simulators like the one used in this study are not available to most rehabilitation specialists, and the technique is seen as having potential, but being somewhat expensive to purchase for small rehabilitation units. An alternative to simulator driving could be to make a structured driving evaluation in real traffic. Further studies should focus on adding more facets to the evaluation, the influence of different raters and settings should be imputed. The next step in the validation of P-Drive should be to modify the assessment tool for on-road evaluations in a sample with differential diagnoses and with different occupational therapists as raters.

ACKNOWLEDGEMENTS

This study was supported by the Postgraduate School of Health Care Sciences at Karolinska Institutet. We thank Karolinska University Hospital, Lars-Olof Ronnevi and Görel Caneman for the generous use of their knowledge and facilities. We also thank Anne G. Fisher for her inspiration and knowledge in instrument development.

REFERENCES

1. Galski T, Ehle HT, McDonald MA, Mackevich J. Evaluating fitness to drive after cerebral injury: basic issues and recommendations for medical and legal communities. J Head Trauma Rehabil 2000; 15: 895–908.
2. Lundqvist A, Gerdle B, Rönnberg J. Neuropsychological aspects of driving after stroke: in the simulator and on the road. App Cogn Psychol 2000; 14: 135–150.
3. Korner-Bitensky NA, Sofer S, Gelinas I, Mazer BL. Brief or new. Evaluating driving potential in persons with stroke: a survey of occupational therapy practices. Am J Occup Ther 1998; 52: 916–919.
4. Rizzo M, McGehee DV, Dawson JD, Anderson SN. Simulated car crashes at intersections in drivers with Alzheimer disease. Alzheimer Dis Assoc Dis 2001; 15: 10–20.
5. Galski T, Ehle HT, Bradley Williams J. Off-road driving evaluations for persons with cerebral injury: A factor analytic study of predriver and simulator testing. Am J Occup Ther 1997; 51: 352–359.
6. Mazer BL, Korner-Bitensky NA, Sofer S. Predicting ability to drive after stroke. Arch Phys Med Rehabil 1998; 79: 743–750.
7. Akinwuntan AE, DeWeerdt W, Feys H, Baten G, Arno P, Kiekens C. Reliability of a road test after stroke. Arch Phys Med Rehabil 2003:1792–1796.
8. Akinwuntan AE, Feys H, DeWeerdt W, Pauwels J, Baten G, Strypstein E. Determinants of driving after stroke. Arch Phys Med Rehabil 2002; 83: 334–341.
9. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989; 70: 308–312.
10. Patomella A, Caneman G, Kottorp A, Tham K. Identifying scale and person response validity of a new assessment of driving ability. Scand J Occup Ther 2004; 11: 70–77.

11. Fisher AG. Uniting practice and theory in an occupational framework: 1998 Eleanor Clarke Slagle lecture. Am J Occup Ther 1998; 52: 509–521.
12. Kielhofner G. Model of human occupation: Theory and application, Third edn. Philadelphia: Lippincott Williams & Wilkins; 2002.
13. Crepeau E. Analysing occupation and activity: a way of thinking about occupational performance, Tenth edition. Philadelphia: Lippincott Williams & Wilkins; 2003.
14. Michon J. A critical view of driver behavioural models: what do we know, what should we do?. New York: Plenium Press; 1985.
15. Linacre JM. Sample size and item calibration stability. Rasch Measurement Transactions 1994; 7: 328.
16. Linacre JM. Winsteps-Rasch Model computer program. Chicago: Winsteps.com; 2005.
17. Fisher AG. The assessment of IADL motor skills: an application of many-faceted Rasch analysis. Am J Occup Ther 1993; 47: 319–329.
18. Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Measurement Transactions 1994; 8: 370.
19. Kottorp A, Bernspang B, Fisher AG. Validity of a performance assessment of activities of daily living for people with developmental disabilities. J Intellect Disabil Res 2003; 47: 597–605.
20. Tham K, Bernspang B, Fisher AG. Development of the assessment of awareness of disability. Scand J Occup Ther 1999; 6: 184–190.
21. Linacre JM. Guidelines for rating scales. In: Midwest Objective Measurement Seminar. Chicago: MESA Research Note #2; 1997.
22. Linacre JM. Detecting multidimensionality: which residual datatype works best? J Outcome Meas 1998; 2: 266–283.
23. Reckase M. Unifactor Latent trait models applied to multifactor tests: results and implications. J Educ Stats 1979; 4: 207–230.
24. Wright BD. Reliability and separation. Rasch Measurement Transactions 1996; 9: 472.
25. Patomella A, Kottorp A. An evaluation of driving ability in a simulator. A good predictor of driving ability after stroke? In: Driving assessment. Portland: Maine, US; 2005.
26. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003; 35: 105–115.
27. Bond T, Fox C. Applying The Rasch Model: fundamental measurement in the human sciences. Mahwah: Lawrence Erlbaum Associates, Inc; 2001.
28. Wright BD, Stone MH. Best test design. Chicago: Mesa Press; 1979.
29. Lundqvist A. Neuropsychological aspects of driving characteristics. Brain Inj 2001; 15: 981–994.
30. Lundqvist A, Rönnberg J. Driving problems and adaptive driving behaviour after brain injury: A qualitative assessment. Neuropsychol Rehabil 2001; 11: 171–185.
31. The National Stroke Registry. Analytic Report for 2003. The County Council of Västerbotten.
32. Körkortshavare efter behörighet [homepage online] Borlänge: Vägverket [updated 2005 March 20]. Availible from: URL: http://www.vv.se/templates/page3wide____4015.aspx
33. Christie N. Assessing driving fitness following brain injury or illness: a research review. TRL Report 1996; 208: 1–21.