

Psychometrics of the Self-Report Safe Driving Behavior Measure for Older Adults

Sherrilene Classen, Pey-Shan Wen, Craig A. Velozo, Michel Bédard,
Sandra M. Winter, Babette Brumback, Desiree N. Lanford

KEY WORDS

- aged
- automobile driving
- caregivers
- psychometrics
- safety
- self report

Sherrilene Classen, PhD, MPH, OTR/L, is Associate Professor, Department of Occupational Therapy, and Director, Institute for Mobility, Activity and Participation, College of Public Health and Health Professions, University of Florida, PO Box 100164, Gainesville, FL 32611-0164; sclassen@phhp.ufl.edu

Pey-Shan Wen, PhD, OTR, is Postdoctoral Fellow, Department of Occupational Therapy, College of Public Health and Health Professions, University of Florida, Gainesville.

Craig A. Velozo, PhD, OTR/L, is Professor, Department of Occupational Therapy, and affiliated with the Institute for Mobility, Activity and Participation, College of Public Health and Health Professions, University of Florida, Gainesville, and is Research Health Scientist at the Rehabilitation Outcomes Research Center, North Florida/South Georgia Veterans Health System, Gainesville.

Michel Bédard, PhD, is Professor, Centre for Research on Safe Driving, Lakehead University, Thunder Bay, ON.

Sandra M. Winter, PhD, OTR/L, is Postdoctoral Associate, Department of Occupational Therapy and Institute for Mobility, Activity and Participation, College of Public Health and Health Professions, University of Florida, Gainesville.

Babette Brumback, PhD, is Associate Professor, Department of Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville.

Desiree N. Lanford, MOT, OTR/L, CDRS, is Certified Driving Rehabilitation Specialist, Department of Occupational Therapy and Institute for Mobility, Activity and Participation, College of Public Health and Health Professions, University of Florida, Gainesville.

We investigated the psychometric properties of the 68-item Safe Driving Behavior Measure (SDBM) with 80 older drivers, 80 caregivers, and 2 evaluators from two sites. Using Rasch analysis, we examined unidimensionality and local dependence; rating scale; item- and person-level psychometrics; and item hierarchy of older drivers, caregivers, and driving evaluators who had completed the SDBM. The evidence suggested the SDBM is unidimensional, but pairs of items showed local dependency. Across the three rater groups, the data showed good person (≥ 3.4) and item (≥ 3.6) separation as well as good person ($\geq .93$) and item reliability ($\geq .92$). Cronbach's α was $\geq .96$, and few items were misfitting. Some of the items did not follow the hypothesized order of item difficulty. The SDBM classified the older drivers into six ability levels, but to fully calibrate the instrument it must be refined in terms of its items (e.g., item exclusion) and then tested among participants of lesser ability.

Classen, S., Wen, P.-S., Velozo, C. A., Bédard, M., Winter, S. M., Brumback, B., & Lanford, D. N. (2012). Psychometrics of the self-report Safe Driving Behavior Measure for older adults. *American Journal of Occupational Therapy, 66*, 233–241. <http://dx.doi.org/10.5014/ajot.2012.001834>

Older driver safety, a public and personal health issue, is best assessed by means of the comprehensive driving evaluation (CDE; American Occupational Therapy Association, 2005; Canadian Association of Occupational Therapists, 2005). The CDE has several limitations, such as the time needed to complete it, high out-of-pocket costs for drivers, exposure to risks, need for trained personnel to administer the test, expensive equipment, liability, limited access, and reporting of unsafe drivers to licensing authorities (Kua, Korner-Bitensky, & Desrosiers, 2007).

Self- and proxy reporting may be a solution to the challenges identified with the CDE. The benefits and limitations of self- and proxy reports have been documented in our previous work (Classen et al., 2010, 2012; Winter et al., 2011). Focusing on the strengths of self- and proxy report measures, we developed the Safe Driving Behavior Measure (SDBM), and research findings have reflected its face and content validity, rater reliability, and rater effects (Classen et al., 2010, 2012; Winter et al., 2011).

The SDBM consists of three sections—A, demographic information; B, driving habits; and C, a driving behavior questionnaire with 68 items (sample items are given in the Supplemental Appendix, available online at <http://ajot.aotapress.net>; navigate to this article, and click on “Supplemental Materials”)—and has a proposed hierarchy of driving tasks that increase in complexity. For example, the instrument indicates that Item 1, “Open car door,” is potentially the easiest item and that Item 68, “Drive on an icy road,” is potentially the most difficult item. On the basis of this principle, one may assume that if a person can drive in “an unfamiliar urban area” (Item 49) without difficulty, then he or she may also be likely to complete the preceding items without difficulty. Understanding the level of safe driving behavior of a participant is a critical step

toward providing an entry point for occupational therapists to plan logical and effective interventions, identify optimal training parameters, and predict future safe driving ability.

The objective of this project was to investigate the item- and person-level psychometrics and item hierarchy of three groups—older drivers, caregivers, and driving evaluators—who had completed the 68-item SDBM. If the SDBM shows reasonable psychometric properties, it will assist occupational therapy generalists with identifying unsafe driving behaviors and provide them with an entry point for delivering preventive services.

Method

This study was approved by the institutional review boards of the University of Florida and Lakehead University.

Participants

We recruited participants in north Florida and Ontario, Canada, by means of advertisements in newspapers, word-of-mouth referrals, and flyers distributed to local community facilities (e.g., retirement communities). A convenience sample of 80 older, community-dwelling, licensed drivers was selected on the basis of the following inclusion criteria: 65–85 yr old, having a valid driver's license, driving at the time of recruitment, having the cognitive ability to complete the SDBM, and having the cognitive and physical ability to participate in an on-road driving test. Participants were excluded if they had been medically advised not to drive, had experienced uncontrolled seizures in the past year, or took medications that caused central nervous system impairment.

Caregivers (18–85 yr old) were included if they were able to report (on the basis of observation) on the older adult's driving behavior. Caregivers were excluded if they showed the presence of physical or mental conditions that impaired the ability to make an active contribution. A total of 80 caregivers were recruited.

At the primary site, the certified driving evaluator, an occupational therapist with 6 yr clinical practice experience, conducted the driving evaluations. At the Canadian site, the driving evaluator was an accredited driving instructor (Province of Ontario) and evaluator with >10 yr of experience.

Procedure

All older drivers and their caregivers gave written informed consent before completing their demographic profiles and the SDBM. Older drivers were tested using our validated clinical battery, which is described in detail elsewhere

(Stav, Justiss, McCarthy, Mann, & Lanford, 2008). Following the protocol, drivers were next evaluated by a trained driving evaluator by means of a standardized on-road driving evaluation (Justiss, Mann, Stav, & Velozo, 2006). The two evaluators (one per site), who were blinded to the participants' SDBM self-ratings or proxy ratings, also completed a SDBM on each driver after the on-road test. Drivers and caregivers received \$50 for their study participation.

Measurement

The SDBM is a 68-item self-report or proxy measure to assess difficulty with the driving task by means of a 5-point adjectival scale ranging from 1 (*cannot do*) to 5 (*not difficult*; Classen et al., 2010). The response option of “not applicable” was used for conditions that some participants experienced (e.g., not driving in snow). For more details on the SDBM, see Classen et al. (2012).

Data Collection and Analysis

All the participant data were collected, stored, and checked in a central secure and password-protected data repository, which was located at the primary site, the University of Florida. We managed the participant demographic data using SPSS Version 17 (SPSS, Inc., Chicago), and we used the rating scale model implemented through the Winsteps Version 3.57 computer program (Linacre, 2005) to conduct Rasch analyses of the rating data. In using the rating scale model, we assumed that the rating scale structure was similar across the 68 items on our instrument. That is, we assumed that the raters used each of the categories of the rating scale in a similar fashion when rating each item (i.e., a 1 on Item 1 was equivalent to a 1 on each of the other items; a 2 on Item 1 was equivalent to a 2 on each of the other items).

We reported only the older drivers' and family members' or caregivers' demographic information and the SDBM's psychometric properties across the three rater groups. Rasch analysis is a one-parameter logistic model that assumes all items have a constant item discrimination parameter. Because of its simplicity, the Rasch model, unlike two-parameter logistic or three-parameter logistic models, does not require large sample sizes to obtain stable estimates and is preferred in the rehabilitation field (Jette & Haley, 2005). For polytomous scales (such as the SDBM), the rating scale model of Rasch analysis, which calibrates the rating scale across all items using the same rating scale structure, is a preferable model for small samples and, hence, adequate to perform data analyses on our sample ($N = 80$ for each

rater group; Linacre, 2000). The measurement model we used is described by the following formula:

$$\text{Log}\left[\frac{P_{nik}}{P_{ni(k-1)}}\right]=B_n - D_i - F_k,$$

where P_{nik} = probability of driver n receiving a rating of k on Item i , $P_{ni(k-1)}$ = the probability of driver n receiving a rating of $k - 1$ on Item i , B_n = the ability of the person, D_i = the difficulty of Item i , and F_k = the difficulty of receiving a rating of k relative to receiving a rating of $k - 1$.

First, we used the principal-components analysis of Rasch-generated residuals (PCAR) to investigate the assumptions of unidimensionality and the correlations of the Rasch-generated residuals to examine the assumption of local independence. We inspected unidimensionality on the basis of eigenvalues and the amount of variance explained by the first component of PCAR. We examined the local dependency on the basis of the strength and pattern of the correlations of the Rasch-generated residuals and evaluated the SDBM rating scale structure, item statistics, person statistics, and item hierarchy.

The rating scale structure was investigated according to three essential criteria: (1) ≥ 10 observations were made per rating category; (2) the average measures (mean of each category) were advanced, indicating that if “cannot do” is -2 logits, the average measure of “a lot of difficulty” should be larger than -2 logits; and (3) the outlier-sensitive mean square fit statistic for each rating scale category should be >2.0 (Linacre, 2002). Item statistics included item difficulty, item fit, item reliability, and item separation, and person statistics included person’s ability, person fit, ceiling and floor effects, person reliability, and person separation. In this article, *person* refers to the older driver.

Item Statistics. *Item difficulty* is an estimate of an item’s underlying difficulty calibrated from the total number of drivers who succeed on the item. *Item fit* was determined by the fit statistics of each item provided by the Winsteps program. The Winsteps program provides two types of fit statistics: information-weighted mean square (infit *MnSq*) and outlier-sensitive mean square (outfit *MnSq*). The driver ratings that a rater assigned in the highest and lowest categories of the scale are weighted less heavily on the infit *MnSq*.

The infit *MnSq* has an expected value of 1. Values >1 signal more variation (i.e., unexplained, unmodeled variation) in a driver’s ratings on the items than expected by the model; values <1 signal less variation in a driver’s ratings on the items than expected by the model. Generally, infit >1 is more of a problem than infit <1 because highly surprising or unexpected ratings that do not fit with the other ratings tend to be more difficult to explain and defend than overly predictable ratings.

By contrast, the outfit *MnSq* statistic is more sensitive than the infit *MnSq* statistic to the occasional highly unexpected and surprising ratings that may occur; therefore, we used outfit statistics. The criteria of the outfit *MnSq* were set from 0.5 to 1.7, and the standardized fit statistics were set from -2 to 2 (Type 1 error rate = 0.05; Wang & Chen, 2005; Wright & Linacre, 1994).

Item reliability represents how well the estimates of the item measures can be replicated when another sample with comparable ability is rated using the same set of items. *Item separation* estimates how well the items are separated by the measured variable.

Person Statistics. *Person’s ability* is an estimate of the driver’s underlying ability and is based on the driver’s performance on a set of items; it is calibrated from the total number of items to which the driver responded successfully. Similar to item fit, *person fit* is determined by the fit statistics of the person; person misfit indicates that one or more of the ratings that the rater assigned to the older driver were surprising or unexpected.

Ceiling effect is defined as $>5\%$ of participants rated at the maximal score, and *floor effect* is defined as $>5\%$ of participants rated at the minimal score. *Person reliability* represents how well the estimate of the driver’s ability can be replicated when other sets of items measuring the same construct are used to rate the same sample of drivers and is analogous to Cronbach’s α with values between 0 and 1. *Person separation index*, measured in standard error units, indicates how well the instrument separates drivers of different levels of safe driving ability. The statistically distinct strata of safe driving ability within the sample of older drivers can be obtained by applying the equation $(4G_p + 1) / 3$, where G_p represents the person separation index (Wright & Masters, 1982). An assessment needs at least two strata to reliably distinguish between safe and unsafe older drivers.

Finally, item hierarchy was evaluated on the basis of the item map provided by the Winsteps program. One of the strengths of the Rasch model is that it can readily handle missing data or “not applicable” answers. That is, the Rasch model does not require a fully crossed rating design; it can easily accommodate partially crossed rating designs that provide sufficient linkage of raters and drivers.

Results

Descriptive Statistics

The demographics for drivers and caregivers are presented in Table 1. It is interesting that although 100% of the caregivers were licensed drivers, only 50% of them drove

Table 1. Demographics and Driving Characteristics of Older Drivers and Their Caregivers

Variable	Older Drivers (<i>n</i> = 80)	Caregivers (<i>n</i> = 80)
Age, yr		
<i>M</i> (<i>SD</i>)	73.26 (5.30)	64.85 (14.03)
Range	65–85	20–85
Gender, <i>n</i> (%)		
Male	41 (51.25)	18 (22.50)
Female	39 (48.75)	62 (77.50)
Race, <i>n</i> (%)		
White	77 (96.25)	79 (98.75)
Other	2 (2.50)	1 (1.25)
Missing data	1 (1.25)	0
Education, <i>n</i> (%)		
College or university	51 (63.75)	39 (48.75)
Vocational training or some college	10 (12.50)	23 (28.75)
Associate's degree	3 (3.75)	3 (3.75)
≤ High school	16 (20.00)	15 (18.75)
Driving history, <i>n</i> (%)		
Drive 7 days/wk	34 (42.50)	40 (50.00)
Drive <7 days/wk	46 (57.50)	40 (50.00)
Licensed driver	80 (100)	80 (100)
Living status, <i>n</i> (%)		
Living alone	25 (31.25)	9 (11.25)
Living with partner or spouse	49 (61.25)	55 (68.75)
Other	6 (7.50)	16 (20.00)
MMSE score, <i>M</i> (<i>SD</i>)	27.67 (1.85)	NA
Number of self-reported medications, <i>M</i> (<i>SD</i>)	6.57 (3.87)	NA

Note. *M* = mean; MMSE = Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975); NA = not applicable; *SD* = standard deviation.

7 days/wk and only 31.3% stated that their independence would be affected if their spouse or partner—the older driver—stopped driving.

Unidimensionality and Local Independency

The PCAR showed the second component had eigenvalues of 4.6, 5.3, and 10.4 for the ratings of the driver, caregiver,

and evaluator groups, respectively. In contrast to the total variances explained by the measure (92.4%, 86.1%, and 90.3%), the second component accounted for only 0.6%, 1.1%, and 1.5% of the total variance for the ratings of the driver, caregiver, and evaluator groups, respectively. The evidence suggested that our measure is unidimensional (Linacre, 2010). For local dependency, the items with high correlations on the Rasch-generated residuals ($r \geq .7$) are shown in Supplemental Table 1, available online at <http://ajot.aotapress.net> (navigate to this article, and click on “Supplemental Materials”). Several pairs of items, the hypothesized “easy items,” showed local dependency.

Rasch Analysis

Rating Scale Structure. Results for rating scale structure indicated the underuse of Category 1, “cannot do.” The observed counts for Category 1 were 26 (driver group), 19 (caregiver group), and 3 (evaluator group). The outfit *MnSq* (perfect outfit *MnSq* = 1) for Category 2, “very difficult,” was 4.86 for the evaluators, indicating that one or more of the ratings that the evaluators assigned in Category 2, for one or more of the items, were quite surprising or unexpected.

Item and Person Statistics. We performed three separate Rasch analyses on the SDBM: for the older drivers, for the caregivers, and for the driving evaluators. The results are summarized in Table 2. In general, the item statistics of Rasch analysis showed 1–13 (1%–19%) misfitting items across three groups of raters; the evaluator group had the highest number of misfitting items. One item showed high infit statistics (misfit) on ratings of both the driver group and the caregiver group: Item 38, “Use a map while driving.” However, it did not show misfit on the ratings of the evaluator group. Instead, the misfitting items on the ratings of evaluator group were Items 1–8 and 10 (hypothesized easiest items); Item 14, “Press gas/

Table 2. Rasch Analysis of the 68-Item Safe Driving Behavior Measure for Driver, Caregiver, and Evaluator

Rater	Driver (<i>n</i> = 80)	Caregiver (<i>n</i> = 80)	Evaluator (<i>n</i> = 80)	Evaluator (<i>n</i> = 79)
Item misfitting, <i>n</i> (%)	2 (3)	1 (1)	13 (19)	8 (12)
Item with minimum estimate value, <i>n</i> (%)	8 (12)	1 (1)	0	3 (4)
Item reliability	.95	.93	.95	.96
Item separation	4.23	3.6	4.19	4.99
Person misfitting, <i>n</i> (%)	6 (8)	6 (8)	6 (8)	6 (8)
Person reliability	0.93	0.92	0.93	0.96
Person separation	3.54	3.49	3.54	4.64
Person strata	5.05	4.99	5.05	6.52
Cronbach's α	.96	.99	.96	.97
Person mean	3.80	4.30	3.80	4.84
Standard deviation (logits)	1.42	1.58	1.42	1.41
Ceiling/floor, <i>n</i> (%)	2 (3)/0	9 (11)/0	0/0	0/0

Note. Column 5 excluded 1 driver, who was identified with high person misfit.

brake”; Item 17, “Emergency brake”; Item 19, “Read sign to react”; and Item 44, “Look before cross” (see Appendix). In addition, good item reliability (>0.93) and good item separation (>3.6) were found across three rater groups. For person (driver) statistics, the results showed 6 (8%) misfitting drivers across the three groups of raters. Good person reliability (>0.92) and good person separation (>3.49) were found across three rater groups. Person means (average of older drivers’ abilities) were about 2 standard deviations higher than the item means across three groups. Additionally, the ratings of the caregiver group showed a slight ceiling effect, because 11% of the drivers rated by their caregivers obtained the maximum score.

Item Hierarchy. We present the item maps rated by evaluator and older driver groups in Figures 1 and 2. These figures demonstrate the older drivers’ abilities and item difficulties on a single linear continuum with equal intervals or logits. Figure 1 (evaluator group) and Figure 2 (older driver group) show that the average of older drivers’ abilities was more than 2 standard deviations higher than the average of item difficulties. The distribution of the older drivers (left side) and items (right side) indicated that this sample had relatively high ability in terms of safe driving behaviors. On the evaluator group’s item map (Figure 1), ratings showed that Item 65, “Drive in a thunderstorm,” was the most challenging item and Item 13, “Reach gas/brake,” and Item 16, “Put in correct gear,” were the easiest items. In comparison, item maps based on the ratings of the driver group (Figure 2) and the caregiver group (not shown) both showed Item 38, “Use a map while driving,” and Item 65, “Drive in a thunderstorm,” as the most difficult items and Item 13, “Reach gas/brake,” as one of the easiest items. Note that the metrics of the item difficulties were different and the midpoints of the scale (zero) are not directly comparable between Figures 1 and 2.

Discussion

We investigated the psychometric properties of the 68-item SDBM by unidimensionality and local independency, rating scale, item- and person-level psychometrics, and item hierarchy across three groups (older drivers, caregivers, driving evaluators).

We tested 80 older drivers, most of whom were White, had a college degree, and were relatively healthy and cognitively intact. Likewise, most of the 80 caregivers were White, but just fewer than one-half of the group had a college degree, and most were female. Although all the caregivers were licensed drivers, one-third stated that their

independence would be affected if their spouse or partner, the older driver, stopped driving. Bias may play a role in the caregivers’ reports in two ways: first, by showing concern for their loved one’s safety, and second, by being more concerned with maintaining their own means of transportation, which was likely to occur in the group that stated that they would be affected if their spouse or partner stopped driving. The existence of this bias has been examined in another article (Classen et al., 2012).

The result of the PCAr was sufficient to assume the SDBM measured a unidimensional construct. The local independency assumption, however, was not supported, especially for the evaluator group’s ratings. The pattern of the residual correlation of the evaluator group’s ratings showed that the hypothesized easy items were highly correlated, which was caused by the low response variances on the easier items on this sample. This result suggests that the easier items (pre-driving items) may be excluded in the final version of the SDBM. The rating scale structure suggested that the “cannot do” category was underused across three rater groups; although it did not discriminate among the ability levels of drivers, it did provide an anchoring point at one end of the rating scale. When we test drivers with ability levels less than that of the current sample, they may very well use the “cannot do” category. However, if future data still indicate underuse, we may have to collapse the “cannot do” and “very difficult” categories.

Item- and person-level psychometrics of the SDBM for each of the three groups revealed incongruence pertaining to (mis)fit. The overlapping misfitting Item 38 in the driver and caregiver groups may need clarification, because group members were not specifically instructed on the type of map (Google map or a geographic positioning system map), which could lead to greater variability in their response choices. Misfitting items (19%) in the evaluator group were problematic. The problem, as determined by post hoc inspection, was the result of the evaluators rating high-ability people as having difficulty with the easiest items. The Rasch model “recognizes” such ratings as inconsistent (misfitting) with the predicted pattern; that is, if people do well on difficult items, they should also do well on easy items. We found one older driver whose response pattern was very different from that of the rest of the older drivers (the person infit statistic was 5; the perfect fit is 1). When this driver was excluded, post hoc analyses revealed that the number of misfitting items for the evaluator group was reduced from 13 to 8 (see Table 2, column 5).

Across the three rater groups, these data displayed good person separation (>3.49) and item separation (>3.6), good item reliability ($>.93$) and person reliability

(>.92), and Cronbach's $\alpha > .96\%$. However, some of the items did not follow the hypothesized order of item difficulty. The evaluator group's ratings showed a different item hierarchy from that of the other two rater groups, potentially because the evaluators wanted to minimize traffic risk and maximize participant safety.

Even though mild ceiling effects existed for the caregivers (11%) and the person mean across the three rater groups was about 2 standard deviations higher than the item mean, this sample was high functioning, so the SDBM may have a sufficient level of challenging items to measure other older adult groups.



Figure 1. The ratings of the evaluator group on older drivers' safe driving ability versus item difficulty.

Note. On the left side of the continuum, each x represents 1 driver, and each # represents 2 drivers, according to the older drivers' abilities, ascending from low, at the bottom, to high, at the top. The items on the right side of the continuum are located on the basis of their difficulty level, ascending from easy, on the bottom, to the most challenging, at the top. The mean, 1 standard deviation, and 2 standard deviations of the older drivers' abilities (left side of the continuum) and of the item difficulties (right side of the continuum) are indicated by M, S, and T, respectively. The Winsteps program anchors the item mean to zero by default. In comparing the left side (person ability) to the right side (item difficulty), person's ability level can be expressed not only by a number but also by a descriptive activity depicted by the parallel item. absn = absence; acr = across; bf = before; convers = conversation; diff = different; fr = from; Ln = lane; Lt = left; Pedestn = pedestrian; Rd = road; Rt = right; traf = traffic; vehi = vehicle.

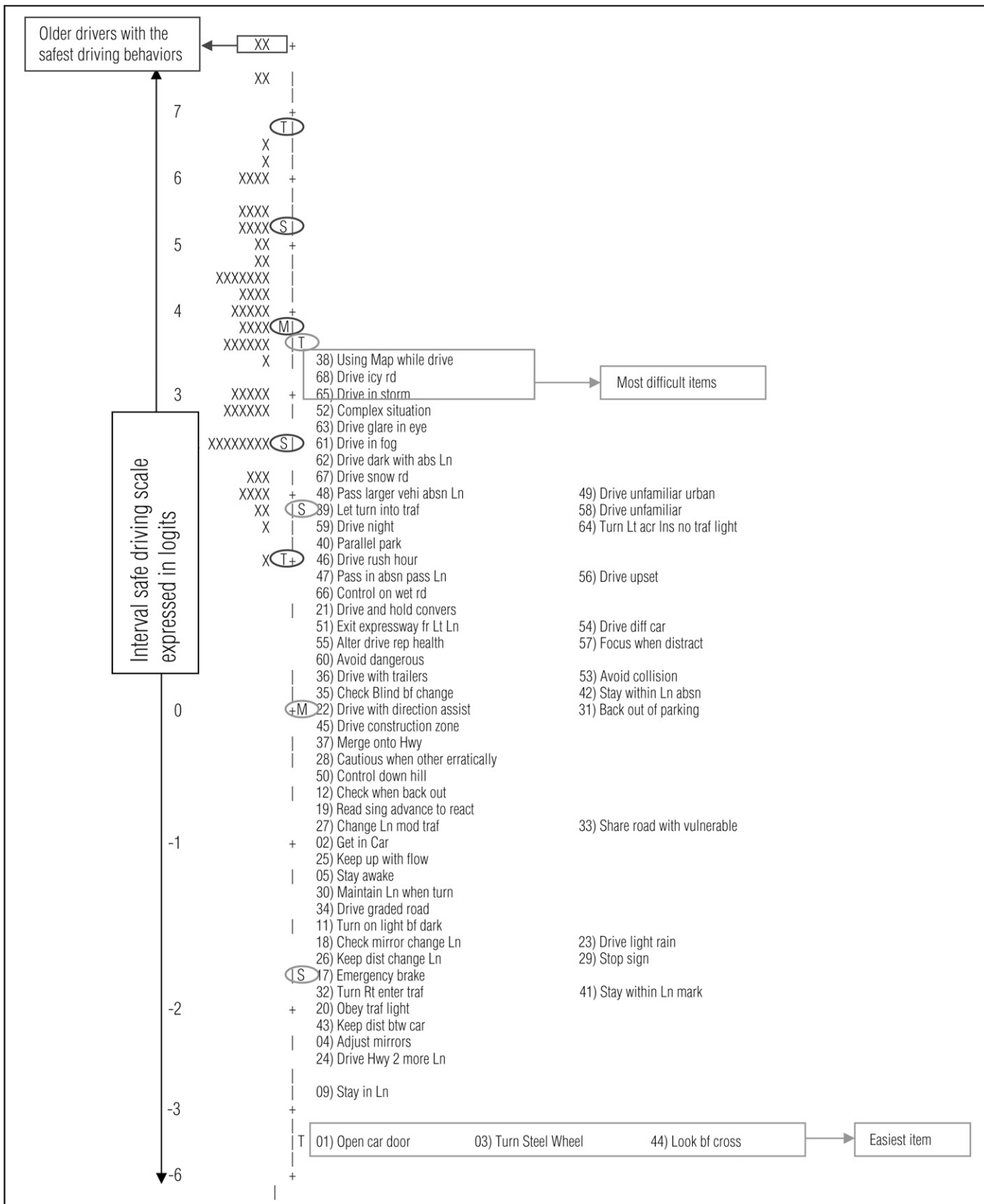


Figure 2. The ratings of the older driver group (self-ratings) on their own safe driving ability versus item difficulty.

Note. On the left side of the continuum, each x represents 1 driver according to the older drivers' abilities, ascending from low, at the bottom, to high, at the top. The items on the right side of the continuum are located on the basis of their difficulty level, ascending from easy, on the bottom, to the most challenging, at the top. The mean, 1 standard deviation, and 2 standard deviations of the older drivers' abilities (left side of the continuum) and of the item difficulties (right side of the continuum) are indicated by M, S, and T, respectively. The Winsteps program anchors the item mean to zero by default. In comparing the left side (person ability) to the right side (item difficulty), person's ability level can be expressed not only by a number but also by a descriptive activity depicted by the parallel item. absn = absence; acr = across; bf = before; convers = conversation; diff = different; fr = from; Ln = lane; Lt = left; Pedestn = pedestrian; Rd = road; Rt = right; traf = traffic; vehi = vehicle.

Implications for Occupational Therapy Practice

Because occupational therapists and driving rehabilitation specialists play an important role in driving evaluation and rehabilitation, it is critical to adopt a psychometrically sound self- or proxy report to identify (un)safe driving behaviors. The SDBM has several advantages over existing measures, as follows:

- The SDBM is a psychometrically sound self- or proxy report measure that is used to assess (un)safe driving behaviors among older adults from a comprehensive person–vehicle–environment approach.
- The SDBM can be completed, using paper and pen, within 20 minutes to minimize respondent burden.
- The items of the SDBM are built on a hierarchy to provide an entry point for occupational therapy assessment and, potentially, intervention.

Limitations and Future Research

This study has several limitations. Caution needs to be exercised when interpreting the data because we can generalize results only to the sample in this study, that is, an educated, mainly White, and cognitively intact group of community-dwelling licensed older drivers. Additionally, several pairs of easy items showed local dependency, and some of them were misfitting as well.

Preliminary findings ($N = 80$; Classen et al., 2012) indicated significant differences between the evaluator and caregiver ratings on 17 items, for which the evaluator rated 7 items more severely than the caregiver and the caregiver rated 10 items more severely than the evaluator. No significant differences were found between the ratings of the evaluator and the driver or between the ratings of the caregiver and the driver. All of these issues must be reconsidered for exclusion if the same pattern holds up after we test drivers with lower ability levels (currently lacking in our sample).

Recently, two driving studies used Item Response Theory to develop or evaluate driving scales: one to convert a standard on-road test to a Rasch scale and the other to develop a measure of driving confidence (Kay, Bundy, & Clemson, 2009; Kay, Bundy, Clemson, & Jolly, 2008; Myers, Paradis, & Blanchard, 2008). Neither studied safe driving from a comprehensive person–vehicle–environment approach, in the driving context, to provide an entry point for occupational therapy intervention. Moreover, the clinical utility of the SDBM is favorable: ≤ 20 min to complete; minimal respondent burden; and items reflecting person, vehicle, and environment domains and driving behaviors requiring very basic to very advanced maneuvers. The instrument can

accurately distinguish people's ability level into five to six strata, and its strengths (i.e., good person and item separation, good item and person reliability, adequate internal consistency, and good clinical utility) motivated us to continue data collection for future analyses.

Conclusion

Our data reflect that the SDBM is efficient and offers the potential to accurately classify a population of older drivers with varying ability levels into distinct groups with more or fewer safe driving behaviors; as such, the SDBM, when tested further and calibrated among drivers with a wide spectrum of ability levels, may provide the first step to identify unsafe driving behaviors and provide occupational therapists with an entry point for delivering preventive services. ▲

Acknowledgment

The project was funded by the National Institute on Aging Grant PAR-06-247 (Principal Investigator, Sherrilene Classen) and University of Florida Center for Multimodal Studies on Congestion Mitigation Grant 00063055 (Principal Investigator, Sherrilene Classen).

References

- American Occupational Therapy Association. (2005). Driving and community mobility. *American Journal of Occupational Therapy, 59*, 666–670. <http://dx.doi.org/10.5014/ajot.59.6.666>
- Canadian Association of Occupational Therapists. (2005). *CAOT position statement: Occupational therapy and driver rehabilitation*. Retrieved August 18, 2008, from www.caot.ca/default.asp?pageid=1353
- Classen, S., Wen, P.-S., Velozo, C. A., Bédard, M., Winter, S. M., Brumback, B., & Lanford, D. N. (2012). Rater reliability and rater effects of the Safe Driving Behavior Measure. *American Journal of Occupational Therapy, 66*, 69–77. <http://dx.doi.org/10.5014/ajot.2012.002261>
- Classen, S., Winter, S. M., Velozo, C. A., Bédard, M., Lanford, D. N., Brumback, B., et al. (2010). Item development and validity testing for a self- and proxy report: The Safe Driving Behavior Measure. *American Journal of Occupational Therapy, 64*, 296–305. <http://dx.doi.org/10.5014/ajot.64.2.296>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198.
- Jette, A. M., & Haley, S. M. (2005). Contemporary measurement techniques for rehabilitation outcomes assessment. *Journal of Rehabilitation Medicine, 37*, 339–345. <http://dx.doi.org/10.1080/16501970500302793>
- Justiss, M., Mann, W., Stav, W., & Velozo, C. (2006). Development of a behind-the-wheel driving performance

- assessment for older adults. *Topics in Geriatric Rehabilitation*, 22, 121–128.
- Kay, L. G., Bundy, A. C., & Clemson, L. M. (2009). Predicting fitness to drive in people with cognitive impairments by using DriveSafe and DriveAware. *Archives of Physical Medicine and Rehabilitation*, 90, 1514–1522. <http://dx.doi.org/10.1016/j.apmr.2009.03.011>
- Kay, L., Bundy, A., Clemson, L., & Jolly, N. (2008). Validity and reliability of the on-road driving assessment with senior drivers. *Accident: Analysis and Prevention*, 40, 751–759. <http://dx.doi.org/10.1016/j.aap.2007.09.012>
- Kua, A., Korner-Bitensky, N., & Desrosiers, J. (2007). Older individuals' perceptions regarding driving: Focus group findings. *Physical and Occupational Therapy in Geriatrics*, 25, 21–40.
- Linacre, J. M. (2000). Comparing “Partial Credit” and “Rating Scale” models. *Rasch Measurement Transactions*, 14, 768.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85–106.
- Linacre, J. M. (2005). WINSTEPS Rasch measurement computer program (Version 3.57) [Computer software]. Chicago: Winsteps.com.
- Linacre, J. M. (2010). *A user's guide to Winsteps/Ministeps Rasch Model computer program*. Beaverton, OR: Winsteps.com.
- Myers, A. M., Paradis, J. A., & Blanchard, R. A. (2008). Conceptualizing and measuring confidence in older drivers: Development of the Day and Night Driving Comfort scales. *Archives of Physical Medicine and Rehabilitation*, 89, 630–640. <http://dx.doi.org/10.1016/j.apmr.2007.09.037>
- Stav, W. B., Justiss, M. D., McCarthy, D. P., Mann, W. C., & Lanford, D. N. (2008). Predictability of clinical assessments for driving performance. *Journal of Safety Research*, 39, 1–7.
- Wang, W.-C., & Chen, C.-T. (2005). Item parameter recovery, standard error estimates, and fit statistics of the WINSTEPS program for the family of Rasch models. *Educational and Psychological Measurement*, 65, 376–404. <http://dx.doi.org/10.1177/0013164404268673>
- Winter, S. M., Classen, S., Bédard, M., Lutz, B., Veloze, C. A., Lanford, D. N., et al. (2011). Focus group findings for a self-report Safe Driving Behavior Measure. *Canadian Journal of Occupational Therapy*, 78, 72–79. <http://dx.doi.org/10.2182/cjot.2011.78.2.2>
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.