
Validity and Critical Driving Errors of On-Road Assessment for Older Drivers

Orit Shechtman, Kezia D. Awadzi, Sherrilene Classen,
Desiree N. Lanford, Yongsung Joo

KEY WORDS

- aged
- automobile driving
- geriatric assessment
- reproducibility of results

OBJECTIVES. We examined the validity of our on-road driving assessment to quantify its outcomes.

METHOD. Older drivers ($N = 127$) completed a driving assessment on a standardized road course. Measurements included demographics, driving errors, and driving test outcomes; a categorical global rating score (pass–fail); and the sum of maneuvers (SMS) score (0–273).

RESULTS. There were significant differences in the SMS ($F = 29.9$, $df = 1$, $p \leq .001$) between drivers who passed the driving test and those who failed. The SMS cutoff value of 230 points was established as the criterion because it yielded the most optimal combination of sensitivity (0.91) and specificity (0.87). The strongest predictors of failure were adjustment to stimuli and lane maintenance errors.

CONCLUSION. The SMS differentiated between passing and failing drivers and can be used to inform clinical decision making.

Shechtman, O., Awadzi, K. D., Classen, S., Lanford, D. N., & Joo, Y. (2010). Validity and critical driving errors of on-road assessment for older drivers. *American Journal of Occupational Therapy*, 64, 242–251.

Orit Shechtman, PhD, OTR/L, is Associate Professor, Department of Occupational Therapy, College of Public Health and Health Professions, and an affiliated researcher with the Institute for Mobility, Activity and Participation and the National Older Driver Research and Training Center, University of Florida, PO Box 100164, University of Florida, Gainesville, FL 32610; oshechtm@phhp.ufl.edu

Kezia D. Awadzi, PhD, is Postdoctoral Associate, Department of Occupational Therapy, College of Public Health and Health Professions, and an affiliated researcher with the National Older Driver Research and Training Center, University of Florida, Gainesville.

Sherrilene Classen, PhD, MPH, OTR/L, is Assistant Professor, Department of Occupational Therapy, College of Public Health and Health Professions; Adjunct Assistant Professor, Department of Epidemiology and Biostatistics; Affiliate Assistant Professor, Department of Behavioral Science and Community Health, College of Public Health and Health Professions; and Director, Institute for Mobility, Activity and Participation and the National Older Driver Research and Training Center, University of Florida, Gainesville.

Desiree N. Lanford, MOT, CDRS, is Staff Occupational Therapist, Department of Occupational Therapy, College of Public Health and Health Professions, and Certified Driving Rehabilitation Specialist, Institute for Mobility, Activity and Participation and National Older Driver Research and Training Center, University of Florida, Gainesville.

Yongsung Joo, PhD, is Assistant Professor, Department of Statistics, Dongguk University, Seoul, Korea.

When occupational therapists make clinical decisions that have a crucial impact on a client's life, they must ensure that their decisions are based on valid assessment instruments. Fitness to drive and sincerity of effort are two areas in which occupational therapists make critical decisions on the basis of assessments with dichotomous results (pass–fail, yes–no). In sincerity-of-effort testing, occupational therapists attempt to determine whether clients are malingering. A person labeled as a *malingerer* stands to lose both financial compensation and employment. Thus, making the decision to classify a client as a malingerer is difficult. Sincerity-of-effort assessments are repeatedly examined for their validity and reliability (Shechtman, 2001; Shechtman, Gutierrez, & Kokendofer, 2005; Shechtman, Hope, & Sindhu, 2007; Shechtman, Sindhu, & Davenport, 2007; Shechtman & Taylor, 2000). Validation of a dichotomous assessment involves sensitivity and specificity analysis to determine whether the test can detect the presence of a condition. Driving assessments also have the potential to have a substantial impact on a person's life and thus should be examined for their validity. It is imperative that occupational therapists know that their clinical decisions are based on valid assessments.

Determining whether a client is fit to drive is difficult because it has life-changing implications for the client. On the one hand, keeping unsafe drivers on the road can put lives and property at risk. On the other hand, revoking a driver's license has negative consequences on the person's independence and engagement in occupation in the areas of work, leisure, and social participation. To make such a life-changing decision, occupational therapists must use valid driving assessments. A standardized assessment instrument, which has the

highest level of measurement validity (Fess, 1995), must undergo rigorous testing to establish its validity and reliability.

Standardization of an assessment instrument includes four steps: (1) developing standardized administration protocols, (2) developing standardized interpretation protocols, (3) determining the reliability and validity values, and (4) establishing its normative values (Fess, 1995). Although some on-the-road driving assessments may have standardized administration and interpretation protocols, most of them do not possess reliability and validity values. Moreover, using the word *standardized* may be confusing because a *standardized driving course* refers to a fixed-route driving course, not to a standardized assessment instrument. *Standardized driving course* actually means that the administration protocol is standardized, not that the actual driving assessment in its entirety is standardized.

Poor driving performance is indicated by driving errors (Blockey & Hartley, 1995; Di Stefano & Macdonald, 2003; Parker, Reason, Manstead, & Stradling, 1995; Staplin, Lococo, McKnight, McKnight, & Odenheimer, 1998). Assessment of driving errors usually occurs during a driving evaluation, which is conducted by a trained driving rehabilitation specialist, usually an occupational therapist. This on-road assessment is considered to be the gold standard for assessing driving safety (Reimer, D'Ambrosio, Coughlin, Kafriksen, & Biederman, 2006) and determining fitness to drive (Di Stefano & Macdonald, 2005; Odenheimer et al., 1994; Shute & Woodhouse, 1990; Yale, Hansotia, Knapp, & Ehrfurth, 2003). However, no single, uniform, widely used, or agreed-on on-road driving assessment exists. Moreover, the outcome of these assessments (passing or failing the driving test) is determined by the subjective judgment of the evaluator, not by a quantifiable driving score (Justiss, Mann, Stav, & Velozo, 2006).

At the National Older Driver Research and Training Center (NODRTC), we have implemented a comprehensive driving evaluation based on experts' opinions obtained through an international consensus conference (Stephens et al., 2005). The on-road driving assessment portion of our comprehensive driving evaluation possesses a standardized road course, a standardized administration protocol, excellent internal consistency of test items (Cronbach's $\alpha = .94$), and excellent reliability (test-retest $r_s = .91-.95$; interrater $r_s = .88-.94$; Justiss et al., 2006). However, this assessment has been only partially validated because it does not possess (1) a well-defined cutoff point for failing the road course, (2) established sensitivity and specificity values, and (3) normative values. Full details of this comprehensive driving evaluation have been published

(Stav, Justiss, McCarthy, Mann, & Lanford, 2008); a detailed description of the administration and interpretation protocols of the NODRTC On-Road Driving Assessment follows.

Administration Protocol

Our standardized road course is a fixed route in Gainesville, FL, with a gradual progression of driving difficulty. The scoring mechanism is based on behavioral errors, which are used to score driving maneuvers. A standardized scoring sheet with a total of 91 driving maneuvers and eight types of possible driving errors is used by driving evaluators while the participant is driving. The driving errors are lane maintenance errors, speed regulation errors, adjustment-to-stimuli errors, yielding errors, signaling errors, vehicle positioning errors, and gap acceptance errors (see Table 1 for the definitions of these errors). For each maneuver, a maximum score of 3 is given if no errors are committed. At the end of the driving course, the sum of the points attained is computed as the *sum of maneuvers score* (SMS), ranging from 0 to 273, with 273 indicating a perfect driving score (zero errors). Ultimately, the driving evaluator determines the global rating score (GRS; passing or failing the test; Justiss et al., 2006; Stav et al., 2008).

Interpretation Protocol

Our test does not have a well-defined cutoff point for failing the road course. The criteria for failing the test are currently based on the driving evaluator's clinical reasoning. This clinical judgment involves the driving evaluator having to intervene (by using the brake or taking hold of the steering wheel), determining that the driver is exhibiting unsafe driving behavior, or both. Unsafe driving behaviors may include consistently drifting to another lane; hitting an object; losing control of the vehicle; not being able to operate the steering wheel, brake, accelerator, or all three; confusing the accelerator and brake pedals; driving through a red light; impeding traffic by driving too slowly; putting other roadway users in jeopardy by driving too fast; and displaying road rage. When a driver is deemed unsafe and the on-road assessment is terminated before the end of the road course, the errors that resulted in the termination of the assessment are called *termination errors*.

Having a quantifiable driving score as a criterion (cutoff value) for failing versus passing our on-road assessment is an important step for establishing its internal validity and providing pass-fail parameters for clinical use. Our on-road assessment has both a categorical

Table 1. Operational Definitions of Driving Errors

Driving Error	Definition
Vehicle positioning (anterior–posterior)	Refers to the position of the vehicle (anterior–posterior) in relation to other vehicles or objects and pavement markings. Captures following distance during forward movement and vehicle spacing during lane changes and merges. Examples of errors include traveling too closely, inadequate space cushion during merge or lane change, and stopping across a crosswalk or too far back from either pavement markings or other vehicles.
Lane maintenance	Refers to the lateral (side-to-side) positioning of the vehicle during driving maneuvers (turns, straight driving, lane changes) and while stopped. Reflects ability to maintain steering control. Examples of errors include drifting out of driving lane, encroachments on perpendicular traffic or wide turns, and parking outside designated space markings.
Speed regulation	Reflects ability to follow and maintain speed regulation limits and having adequate control of the vehicle's acceleration and braking features. Examples of errors include not coming to a complete stop at a stop sign, traveling too slow or too fast, inadequate merging speed regulation, and abrupt or inappropriate braking or acceleration.
Yielding	Refers to giving right of way when appropriate. Refers to the ability to recognize common rules of road safety. Yielding is assessed at four-way or two-way stop intersections, right turns on red, and merges.
Signaling	Refers to proper use of turn signals. Examples of errors include leaving the turn signal on, not using the turn signal when turning, and using the turn signal inappropriately (wrong signal for given turn, signaling too short until maneuver).
Adjustment to stimuli or traffic signs	Reflects the ability to appropriately respond to driving situations. Captures the ability to adjust appropriately to changing road sign information, other vehicle movements, and pedestrian movements and the ability to recognize potential hazards. Examples of errors include not adjusting speed regulation for posted limits, not following proper evaluator instructions, choosing improper lane from posted signage, and improper response to traffic or pedestrian movement.
Gap acceptance	Refers to choosing an appropriately safe time or spacing distance to cross in front of oncoming traffic (unprotected left turn). Errors in gap acceptance are based on evaluator judgment given the speed regulation of oncoming traffic and number of lanes to be crossed. Errors in gap acceptance consist of driver estimates that are both too short and too long for the given speed regulation and distance to be traveled.

outcome (GRS) and a numerical outcome (SMS). The relationship between these two outcomes can be examined using sensitivity and specificity analysis and the receiver operating characteristic (ROC) curve analysis. These statistical analyses examine whether forming categories of pass and fail (GRS) corresponds to a particular point value of the SMS (which is based on the number of driving errors committed by the driver).

The internal validity of our on-road evaluation can be assessed using sensitivity and specificity analysis, which is commonly used to determine whether a test can detect the presence or absence of a condition (Shechtman, 2001). *Sensitivity* is defined as the test's ability to obtain a positive test when the condition really exists (a true positive), and *specificity* is defined as the test's ability to obtain a negative test when the condition is really absent (a true negative; Portney & Watkins, 2000). For the on-road driving assessment, a *positive test* means that the person failed the driving test. Any value within the range of the SMS (0–273) may be selected as the cutoff value, below which the driving test is considered positive. However, the number of false positives (those who receive a failing score but pass the road test) and false negatives (those who receive a passing score but fail the road test), and thus the sensitivity and specificity values, changes with the cutoff value (Portney & Watkins, 2000; Shechtman,

2000). The larger the SMS cutoff value is, the greater the sensitivity and the smaller the specificity are, and vice versa.

A way of examining the effects of applying different cutoff values to two overlapping distributions of scores is to plot the ROC curve (McNicol, 1972). The ROC curve is a plot of the rate of true positives (hits; sensitivity) against the rate of false positives (misses; $1 - \text{specificity}$) resulting from application of many arbitrarily chosen cutoff points of SMS. Therefore, the ROC curve demonstrates the effectiveness of using different cutoff values and reveals the optimal SMS cutoff value.

When taking a driving test, driving errors contribute to failing the test. However, some driving errors are less critical than others; in other words, certain errors may be better tolerated so that a driver could pass the test despite making a few slight errors. For example, committing a signaling error may be less detrimental than making a lane maintenance error. Exactly which type of errors or how many of them result in failing a driving test is unknown. Staplin et al. (1998) suggested that among older adults, the driving errors that are strongly predictive of crashes include lane change with an unsafe gap, failure to stop completely at a stop sign, stopping over a stop bar, improper turning path, and stopping for no reason. Studies examining driving assessment of drivers with

dementia have supported this suggestion (Hunt, Morris, Edwards, & Wilson, 1993; Hunt et al., 1997).

The purpose of this study was to examine the internal validity of our on-road driving assessment and to quantify its outcome. Our specific aims were (1) to examine whether the numerical SMS is able to differentiate between drivers who pass the driving test and those who fail it, (2) to establish a criterion SMS for passing or failing the driving test by finding the optimal combination of sensitivity and specificity values, and (3) to discern which types of driving errors are most predictive of failing the driving test.

Method

Participants

We used a convenience sample of 127 volunteers, ≥ 65 years old (average age = 74.9, standard deviation = 6.4) with a valid driver's license. These older adults lived in the community and had a variety of comorbidities but had to have been seizure free for the past year. A detailed description of the participants' health conditions is available elsewhere (Classen et al., 2008). The study was approved by the University of Florida, Gainesville, Institutional Review Board, and all participants were provided written informed consent.

Setting

The participants completed an on-road driving assessment on a standardized road course. The standardized road course is currently used to test driver performance at the NODRTC in Gainesville, FL. The details of the standardized course and testing are described elsewhere in detail (Justiss et al., 2006; Stav et al., 2008).

Measures

Independent variables included the demographic variables of gender and age. Dependent variables included the GRS (pass or fail), the SMS (0–273 points, with 273 indicating perfect driving), and eight types of driving errors (Table

1), which were recorded by the driving rehabilitation specialists as participants were driving the standardized road course.

Procedures

The in-vehicle, on-the-road driving assessment was administered by the driving rehabilitation specialists (who were also registered occupational therapists; one of the evaluators was author Desiree Lanford). The driving evaluator sat in the passenger seat of the test vehicle, a 2004 Buick Century equipped with a dual brake. The interrater reliability among the driving evaluators was good to excellent (intraclass correlation coefficients = 0.80–1.00; Posse, McCarthy, & Mann, 2006). We used an open, standardized road course in Gainesville, FL, with varying levels of complexity, including residential driving and highway driving. The driving performance data included driving errors, an SMS (sum of all errors), and a GRS. The GRS has four outcomes: pass, pass with recommendation, fail with recommendation, and fail. For this study, we condensed both pass categories into the pass outcome and both failing categories into the fail outcome.

Data Analysis

We used descriptive statistics to determine the age and SMS of men and women who passed and failed the driving test and the frequencies of errors. Inferential statistics included a 2×2 analysis of variance (ANOVA; Gender \times GRS) to determine whether SMS differences existed between men and women who passed or failed the driving test. Sensitivity and specificity values were calculated for the SMS cutoff values of 0–273 (see Figure 1 for an example of calculating sensitivity and specificity). We calculated the overall error rate by using the equation $(1 - \text{sensitivity}) + (1 - \text{specificity})$. To find the optimal cutoff value of SMS, we generated a ROC curve on the basis of multiple SMS cutoff values. The ROC curve was

		GRS	
		+	–
		(Fail)	(Pass)
SMS	+	a True Positives (Hits; $n = 21$)	b False Positives (Misses; $n = 14$)
	–	c False Negatives (False Alarms; $n = 2$)	d True Negatives (Correct Rejections; $n = 90$)

Figure 1. Calculating sensitivity and specificity for the SMS 230-point cutoff value: sensitivity = $a/(a + c)$, or $21/(21 + 2) = 0.91$; specificity = $d/(b + d)$, or $90/(14 + 90) = 0.87$.

Note. SMS = sum of maneuvers score; GRS = global rating score.

generated by plotting the true-positive rate (hits) against the false-positive rate (misses) for each of the cutoff values. The false-positive rate was calculated by subtracting the specificity value from 1.00 (1.00 – specificity). To obtain an index of discriminability, we calculated the proportional area under the curve (McNicol, 1972).

To examine which driving errors are most predictive of poor performance on the driving test, we performed two types of regression analysis. We used logistic regression to determine which errors are predictive of the GRS because this variable is categorical. We used forward stepwise regression to determine which errors are predictive of the SMS. Statistical analyses were performed using SPSS 15 (SPSS, Inc., Chicago).

Results

Descriptive and Inferential Statistics

Overall, 23 of 127 participants (18.1%) failed the driving test, which consisted of 12 of 68 men (17.6%) and 11 of 59 women (18.6%). The mean and standard deviation of the SMS was 239 ± 25 (see Table 2 for complete descriptive statistics). The two-way ANOVA (Gender × GRS) revealed significant differences in SMS between drivers who passed the test and those who failed it ($F = 29.9$, $df = 1$, $p \leq .001$) but no significant gender differences in SMS.

Sensitivity and Specificity Analysis

To examine whether the GRS (pass–fail) corresponded to the SMS, we constructed a ROC curve on the basis of on multiple sensitivity and specificity values encompassing the entire range of SMS. The ROC curve revealed that the SMS cutoff value of 230 yielded the most optimal combination of sensitivity (0.91) and specificity (0.87), that is, the combination yielding the lowest overall error rate (22%; Table 3). The proportional area under the ROC curve was 90.6% (Figure 2).

Regression Analysis

The odds ratios from the logistic regression for the GRS are provided in Table 4. Note that in a logistic regression, odds ratios are the exponent of the regression parameters.

Therefore, if an odds ratio is significantly different from 1, the corresponding regression parameter is significantly different from 0. Only when the estimates of the odds ratios are significant at the 95% confidence interval can we predict the probability of failing a driving test using these estimates. For example, when the adjustment-to-stimuli error increases as much as 1, then the odds of failing the driving test is predicted to increase by a factor of 2.25 (Table 4). The strongest predictor of failing the driving test was adjustment-to-stimuli errors, followed by lane maintenance errors. These two errors were the only significant predictors of failing the on-road driving test. All the rest of the driving errors, as well as age and gender, did not predict failing the driving test (Table 4).

The stepwise regression analysis for the SMS is reported in Table 5. The strongest predictor of low test score was committing lane maintenance errors, which accounted for 58% of the variance ($R^2 = .58$). The other significant errors that contributed to the variance, in order of contribution, were speed regulation errors, adjustment-to-stimuli errors, yielding errors, and signaling errors (Table 6). The nonsignificant predictors were vehicle positioning errors and gap acceptance errors. Table 6 summarizes the type of driving error, its rank of contribution based on the stepwise regression, the number of drivers who committed these types of errors, and the total number of errors committed in each category (some drivers committed multiple errors of the same type when going through the 91 driving maneuvers).

Discussion

Currently, most on-road driving assessments have only a pass–fail outcome that is based on driving evaluators’ clinical reasoning and not on a quantifiable, numerical test score. A standardized on-road driving assessment with a quantifiable score would allow for greater objectivity in determining whether a driver is fit to drive. Our on-road driving assessment has two outcome measures: the categorical GRS, which is based on clinical judgment, and the numerical SMS, which is based on the number of driving errors. This study’s findings indicate that senior drivers (\geq age 65) who failed the on-road

Table 2. Age and Sum of Maneuver Score (SMS) of Men and Women Who Passed and Failed the Driving Test

Variable	Total			Men			Women		
	All	Pass	Fail	All	Pass	Fail	All	Pass	Fail
<i>n</i> (%)	127	104 (81.9)	23 (18.1)	68 (53.5)	56 (82.4)	12 (17.6)	59 (46.5)	48 (81.4)	11 (18.6)
Age (years; <i>M</i> ± <i>SD</i>)	74.9 ± 6.4	73.9 ± 6.0	78.7 ± 6.6	74.85 ± 6.7	73.86 ± 6.2	79.50 ± 7.2	74.71 ± 6.1	73.92 ± 5.9	78.18 ± 6.1
SMS (<i>M</i> ± <i>SD</i>)	239 ± 25	246 ± 14.5	205 ± 33.3	240 ± 24	247 ± 14.1	209 ± 35.4	238 ± 26	246 ± 15.1	200 ± 31.9

Note. *M* = mean; *SD* = standard deviation.

Table 3. Sensitivity and Specificity Values for Specific Sum of Maneuvers Score (SMS) Cutoff Values in Determining Global Rating Score (Pass-Fail)

SMS	Sensitivity	Specificity	Total Error ^a
112	0.04	1	0.96
122	0.04	1	0.96
132	0.04	1	0.96
142	0.09	1	0.91
152	0.09	1	0.91
162	0.09	1	0.91
172	0.13	1	0.87
182	0.13	1	0.87
192	0.3	1	0.7
202	0.3	0.99	0.71
212	0.57	0.98	0.45
222	0.78	0.93	0.29
228	0.87	0.89	0.24
230	0.91	0.87	0.22
232	0.91	0.84	0.25
242	0.91	0.68	0.41
252	0.96	0.36	0.68
262	0.96	0.09	0.95
272	1	0.04	0.96
273	1	0	1

^aTotal error = (1 - sensitivity) + (1 - specificity).

assessment (according to the categorical GRS) had a significantly lower SMS than drivers who passed the test. Thus, the SMS can indeed differentiate between older drivers who failed the test and those who passed it, which makes this quantifiable score a valid measure of determining fitness to drive. To be valid, however, a dichotomous

(pass-fail) assessment must also have sufficient sensitivity and specificity values (Shechtman, 2001; Shechtman et al., 2005; Shechtman, Hope, et al., 2007; Shechtman, Sindhu, & Davenport, 2007; Shechtman & Taylor, 2000).

Using sensitivity and specificity as well as ROC curve analyses, we established the SMS of 230 points as the optimal criterion score for passing our on-road assessment. Clinically, this criterion score indicates that people who receive <230 points would fail our driving test. When using this cutoff value as the criterion for determining pass-fail, we found that 9% (1 - sensitivity) of older drivers who failed the test had ≥230 points (false negatives) and that 13% (1 - specificity) of older drivers who passed the test had <230 points (false positives). Choosing a different cutoff point would result in a decrease in either sensitivity or specificity because of the trade-offs between them (as sensitivity increases, specificity decreases, and vice versa). These trade-offs depend on how stringent the criterion is (Portney & Watkins, 2000). Using a smaller cutoff value (a less stringent criterion) would decrease sensitivity (a Type 1 error), thereby increasing the clinician's chance of wrongly classifying an unsafe driver as fit to drive. Conversely, using a greater cutoff value (a more stringent criterion) would decrease specificity (a Type 2 error), thereby increasing the chance of erroneously identifying a safe driver as unfit to drive. Clinicians who use driving assessments need to make a compromise of sorts by deciding what type of error they are willing to make.

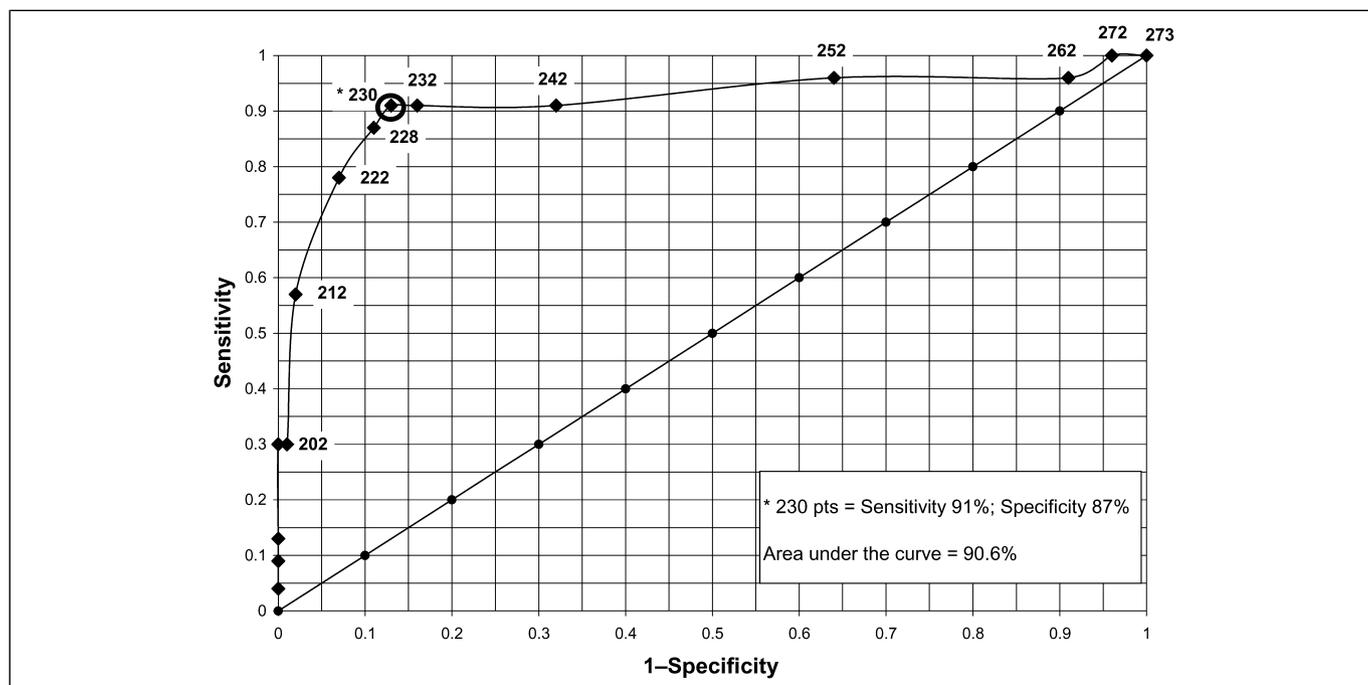


Figure 2. Receiver operating characteristic curve of multiple sum of maneuvers score cutoff points (diamonds).

Note. The optimal cutoff value is 230 points (circled). The dots (diagonal line) represent values that are no better than chance.

Table 4. Logistic Regression Results: Types of Driving Errors as Predictors of Failing the Driving Test

Parameter	Estimate	<i>p</i>	Odds Ratio	Lower Confidence Limit	Upper Confidence Limit
Age	0.1001	.0896	1.105	0.985	1.241
Gender (male)	-0.6681	.3814	0.513	0.115	2.289
Vehicle positioning	-0.01	.9311	0.99	0.789	1.243
Speed regulation	0.0019	.9559	1.002	0.936	1.072
Lane maintenance	0.0998	.0284*	1.105	1.011	1.208
Signaling	-0.0834	.2662	0.92	0.794	1.066
Yielding	-0.1917	.7003	0.826	0.311	2.191
Adjustment to stimuli	0.8133	.0036*	2.255	1.305	3.898
Gap acceptance	1.2056	.169	3.339	0.599	18.605

**p* ≤ .05.

According to Portney and Watkins (2000), clinicians who use a screening tool should decide what levels of sensitivity and specificity are acceptable, based on the consequences of false negatives and false positives. In the case of a life-threatening disease, sensitivity is more important because a misdiagnosis may prove to be fatal (Portney & Watkins, 2000). Conversely, specificity is more important when the harm caused by diagnosing a condition when it does not exist is high. An example is sincerity of effort: When misdiagnosing a sincere client as a malingerer, clinicians risk negatively affecting the person's future treatment, income, and work (Shechtman, 2001; Shechtman & Taylor, 2000). Therapists who use driving assessments to determine fitness to drive must decide what levels of sensitivity and specificity are acceptable. The consequences of low sensitivity or low specificity are mistakenly identifying drivers as safe or unsafe, a clinical mistake that may have a serious impact on the client.

Low sensitivity (a Type 1 error) entails mistakenly identifying unsafe drivers as fit to drive. The clinical consequences of this type of error (allowing unsafe drivers to continue driving) involve risking the life, injury, and

property of the driver and others. In the current study, 9% of drivers who failed the test had ≥230 points; if we based the driving test results only on the cutoff value, we would have allowed these 2 of 23 unsafe drivers to keep driving. Conversely, low specificity (a Type 2 error) entails mistakenly classifying safe drivers as unfit to drive. This clinical mistake may result in revoking clients' driver's license when they are still fit to drive. The consequences of this type of error involve negative effects on clients' independence and quality of life. Driving cessation could have a far-reaching impact on the older adult's engagement in occupation in the areas of work, leisure, and social participation. In the current study, 13% of drivers who passed the test had <230 points; if we based the driving test results only on the cutoff value, we would have disallowed 14 of 104 safe drivers from continuing to drive.

The reason that no absolute agreement exists between the GRS and the SMS is that some driving errors are more critical than others. Thus, drivers with a higher SMS might fail the driving test as a result of making a critical error (such as driving through a red light), whereas drivers with a lower SMS may pass the test because they did not make critical driving errors. Because differences in weight exist between the type of errors, we used regression analysis to examine which type of driving errors are more predictive of failing the on-road driving assessment.

The logistic regression between the GRS and the type of errors showed that only two types of errors were significantly related to failing our on-road driving test. A driver had a twofold probability of failing the test if he or she made an adjustment-to-stimuli error and only a 10% higher probability of failing the test if he or she committed a lane maintenance error. Making any other error was not significantly related to passing or failing the test. The results of the logistic regression—that age and gender did not predict failing the driving test—were in agreement with the ANOVA results, which revealed no significant gender differences in failing the test.

Table 5. Model Summary of Forward Stepwise Regression Analysis for Driving Errors Predictive of Failing the Driving Test

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Standard Error of the Estimate	Change Statistics			
					<i>R</i> ² Change	<i>F</i> Change	Degrees of Freedom	<i>p</i>
1	.763 ^a	.582	.579	16.238	.582	174.250	1, 125	.000
2	.848 ^b	.719	.714	13.381	.136	60.074	1, 124	.000
3	.866 ^c	.750	.744	12.660	.032	15.522	1, 123	.000
4	.871 ^d	.759	.751	12.494	.008	4.288	1, 122	.040
5	.878 ^e	.770	.761	12.241	.012	6.112	1, 121	.015

^aPredictors: (Constant), lane maintenance errors.

^bPredictors: (constant), lane maintenance errors, speed regulation errors.

^cPredictors: (constant), lane maintenance errors, speed regulation errors, adjustment to stimuli errors.

^dPredictors: (constant), lane maintenance errors, speed regulation errors, adjustment to stimuli errors, yielding errors.

^ePredictors: (constant), lane maintenance errors, speed regulation errors, adjustment to stimuli errors, yielding errors, signaling errors.

Table 6. Ranking of the Type of Driving Errors Predictive of Failing the Driving Test (Based on Stepwise Regression), the Number of Drivers Committing Each Error, and the Total Number of Errors Committed in Each Category

Type of Error	Rank	No. of Drivers (<i>N</i> = 127)	Total No. of Errors
Lane maintenance	1	123	1,539
Speed regulation	2	121	1,818
Adjustment to stimuli	3	63	186
Yielding	4	30	38
Signaling	5	110	591
Vehicle positioning	6	117	577
Gap acceptance	8	14	16

Note. Some drivers committed multiple errors of the same type when going through the 91 driving maneuvers.

The stepwise linear regression between the SMS and the types of errors revealed that several errors are predictive of failing the on-road assessment. The single strongest predictor of failing the test was committing a lane maintenance error, which accounted for 58% of the variance. A model that included all significant types of errors (lane maintenance errors, speed regulation errors, adjustment-to-stimuli errors, yielding errors, and signaling errors) accounted for 77% of the variance (Table 3). The non-significant predictors included two types of errors: vehicle positioning errors and gap acceptance errors. The clinical applications of our findings may be used by occupational therapists and certified driving rehabilitation specialists for both evaluation and intervention. On the basis of our results, therapists may need to pay more attention to the statistically significant types of errors during a driving assessment, put more focus on remediating these types of errors during intervention, or both.

Although the stepwise regression identified more errors as predictors of driving performance than did the logistic regression, they both identified as critical two types of errors: lane maintenance errors and adjustment-to-stimuli errors. Thus, these errors are predictive of both outcome scores of our on-road driving assessment: the categorical GRS and the numerical SMS. It is not surprising that total agreement between the two regression analyses does not exist because the two outcome measures are not identical, as suggested by the 22% overall error rate in identifying pass–fail using the numerical SMS.

Speed regulation errors, which were the second strongest predictor of the SMS, did not significantly predict the GRS (passing or failing the test). A possible explanation is that our database did not differentiate between driving too slowly and driving too fast when recording speed regulation errors. It is probable that most of the older drivers who received a speed regulation error

drove too slowly, but not so slow that they were deemed unsafe enough to cause them to fail the driving test. Thus, driving too slowly could have reduced their SMS without affecting their GRS.

The convenience sampling was a limitation of this study because more drivers passed the test than failed it. In addition, although we identified the types of errors that predict performance on the driving test, we still do not know whether these errors could predict future crashes or citations. To be ecologically valid, driving assessments must predict actual driving performance in the community. Similarly, we want to clarify that although we have performed an important and necessary step in assessment validation (i.e., assessing its internal validity), the tool’s actual effectiveness can only be determined by testing ecological validity. Still, incorporating the SMS with the subjective overall pass–fail clinical judgment (GRS) partially ameliorates the problem of subjectivity and strengthens our assessment in comparison with other existing assessments that use only pass–fail scores. Thus, despite the study’s limitations, it provides important information regarding the validity of our on-road driving assessment.

The immediate clinical applications of this study involve the driving evaluators at the NODRTC, who now use the study’s findings in their day-to-day operation. On the basis of these findings, we constructed a decision tree (Figure 3) to determine whether the GRS is valid and to increase evaluators’ confidence that their clinical decision is correct. The decision tree is based on two factors and thus has two levels: (1) the relationship between the cutoff value and the individual driver’s SMS and (2) the type of errors committed by the driver. Specifically, when determining that an older driver passed the on-road assessment, the evaluator would first check whether the SMS falls above the cutoff value (230): If it does, then the pass score is confirmed; if it does not, then the evaluator would check to see whether the driver committed a critical error. If she or he did not, then the pass score stands; if she or he did, then the driver would be retested (Figure 3). If on retest, the decision tree process results in yet another retest, then the driver must undergo remediation before being tested for the third time. A similar decision tree would be followed for a fail score (see Figure 3 for the decision tree). Critical errors include termination errors (which result in the evaluator terminating the road test) and errors that were found to significantly predict the GRS, namely adjustment-to-stimuli errors and lane maintenance errors.

The far-reaching applications of this study offer a model for clinician–researcher collaboration in the area

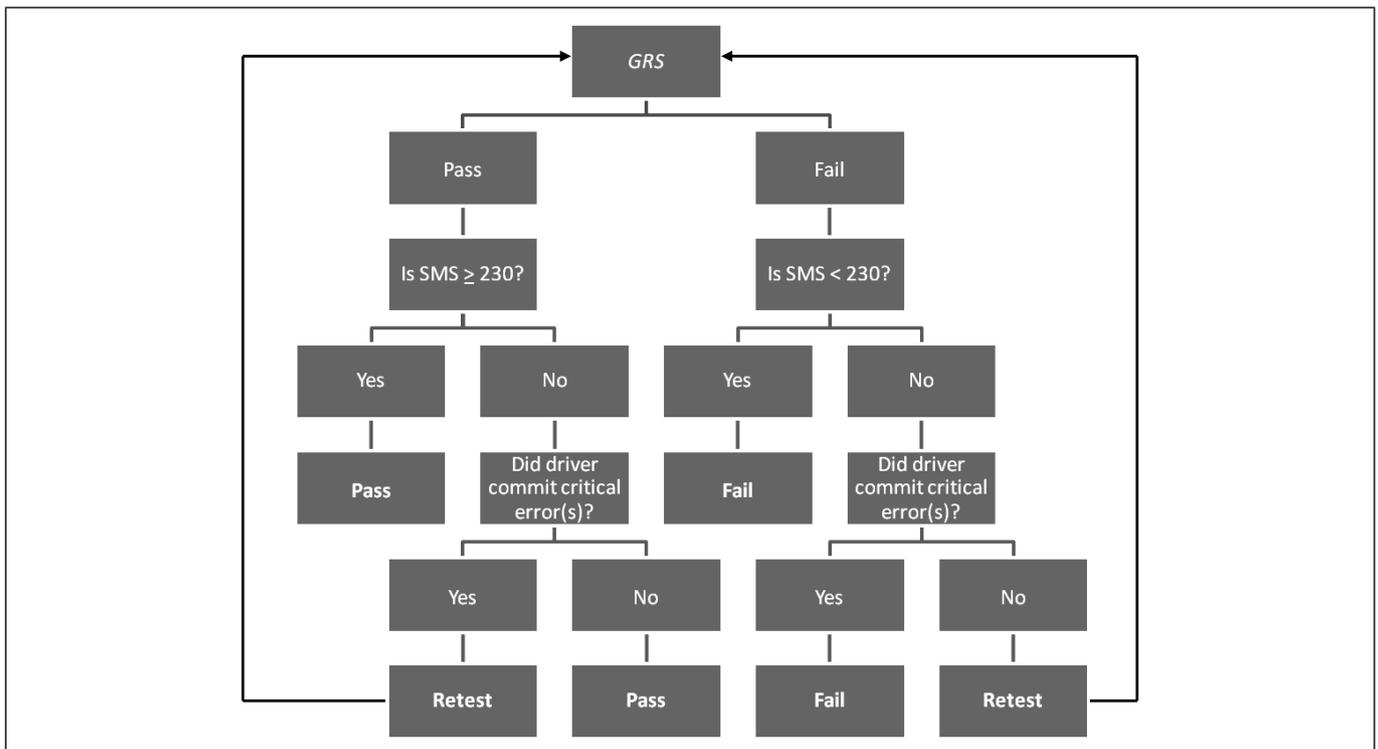


Figure 3. Decision tree for confirming the global rating score (GRS; pass-fail).

Note. Decision boxes are shaded. SMS = sum of maneuvers score; critical errors = termination errors and predictive errors (adjustment-to-stimuli errors, lane maintenance errors, or both).

of driving (and possibly in other occupational therapy areas). In this model, practice informs research to begin with; in other words, clinicians first use an assessment and then researchers test its validity by using statistical methodology (Sensitivity and Specificity, ROC). Then, research informs practice: Clinicians use the results of the research study to modify the assessment. Using this model could improve clinicians' confidence in correctly evaluating clients and helps to reduce errors in clinical judgment.

Finally, the proportional area under the ROC curve, a measure of discriminability (i.e., the ability to discriminate between passing and failing the driving test), was 90.6% (Figure 2). The area under the ROC curve is an index of the degree of separation (or overlap) between the distributions of true positives (signal) and false positives (noise; McNicol, 1972). A perfect diagnostic test has an area of 100% (McNicol, 1972). A larger area under the curve indicates a better ability to discriminate between failing and passing the test. Our finding that the area under the ROC curve was >90% is an additional indication of the ability of our scoring system to discriminate between passing and failing the on-road driving assessment.

Conclusions

This study represents an essential step in establishing the psychometric properties and strengthening the clinical

utility of an assessment tool in an important occupational therapy practice area of older adult fitness to drive. We assessed the SMS's effectiveness to serve as a quantifiable and objective method of determining passing or failing the road test. Testing the SMS against the GRS (both of which are derived from the same assessment) allowed us to establish our assessment's internal validity. We found that the numerical outcome measure of SMS can differentiate between older drivers who failed the test and those who passed it ($F = 29.9$, $df = 1$, $p \leq .001$; sensitivity = 0.91; specificity = 0.87; overall error rate = 22%). Using this numeric outcome provides us with a standardized interpretation protocol and increases the objectivity of our assessment. In addition, we identified the most critical errors in predicting both the categorical and the numerical outcomes of the test.

These parameters provided useful information for creating a decision tree to inform clinical decision making among driving evaluators. The decision tree combines the cutoff value for passing or failing the driving test (objective assessment) with the knowledge of which driving errors are most significant in failing that test, which allows occupational therapists and certified driving rehabilitation specialists to determine fitness to drive with greater confidence. The findings also offer a clinician-researcher collaboration model and create possible research opportunities to examine the

ecological validity of on-road driving assessments in identifying unsafe drivers in the community (e.g., when driving their own vehicle without the presence of a driving evaluator). ▲

Acknowledgment

We acknowledge the University of Florida's College of Public Health and Health Professions for funding this project; the Gainesville Traffic Engineering Department, Gainesville, FL; and the National Older Driver Research and Training Center (NODRTC), University of Florida, Gainesville.

References

- Blockey, P. N., & Hartley, L. R. (1995). Aberrant driving behaviour: Errors and violations. *Ergonomics*, *38*, 1759–1771. doi:10.1080/00140139508925225
- Classen, S., Horgas, A., Awadzi, K., Messinger-Rapport, B., Shechtman, O., & Joo, Y. (2008). Clinical predictors of older driver performance on a standardized road test. *Traffic Injury Prevention*, *9*, 456–462. doi:10.1080/15389580802260026
- Di Stefano, M., & Macdonald, W. (2003). Assessment of older drivers: Relationships among on-road errors, medical conditions and test outcome. *Journal of Safety Research*, *34*, 415–429. doi:10.1016/j.jsr.2003.09.010
- Di Stefano, M., & Macdonald, W. (2005). On-the-road evaluation of driving performance. In J. M. Pellerito (Ed.), *Driver rehabilitation and community principles and practice* (pp. 255–274). St. Louis, MO: Elsevier/Mosby.
- Fess, E. E. (1995). Guidelines for evaluating assessment instruments. *Journal of Hand Therapy*, *8*, 144–148.
- Hunt, L. A., Morris, J. C., Edwards, D. F., & Wilson, B. S. (1993). Driving performance in persons with mild senile dementia of the Alzheimer type. *Journal of the American Geriatrics Society*, *41*, 747–752.
- Hunt, L. A., Murphy, C. F., Carr, D., Duchek, J. M., Buckles, V., & Morris, J. C. (1997). Reliability of the Washington University Road Test: A performance-based assessment for drivers with dementia of the Alzheimer type. *Archives of Neurology*, *54*, 707–712.
- 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.
- McNicol, D. (1972). *Primer of signal detection theory*. London: Allen & Unwin.
- Odenheimer, G. L., Beaudet, M., Jette, A. M., Albert, M. S., Grande, L., & Minaker, K. L. (1994). Performance-based driving evaluation of the elderly driver: Safety, reliability, and validity. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, *49A*, 153–159.
- Parker, D., Reason, J. T., Manstead, A. S. R., & Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics*, *38*, 1036–1048. doi:10.1080/00140139508925170
- Portney, L., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall Health.
- Posse, C., McCarthy, D., & Mann, W. (2006). A pilot study of interrater reliability of the Assessment of Driving-Related Skills: Older driver screening tool. *Topics in Geriatric Rehabilitation*, *22*, 113–120.
- Reimer, B., D'Ambrosio, L. A., Coughlin, J. E., Kafrissen, M. E., & Biederman, J. (2006). Using self-reported data to assess the validity of driving simulation data. *Behavior Research Methods*, *38*, 314–324.
- Shechtman, O. (2000). Using the coefficient of variation to detect sincerity of effort of grip strength: A literature review. *Journal of Hand Therapy*, *13*, 25–32.
- Shechtman, O. (2001). The coefficient of variation as a measure of sincerity of effort of grip strength, Part II: Sensitivity and specificity. *Journal of Hand Therapy*, *14*, 188–194.
- Shechtman, O., Gutierrez, Z., & Kokendofer, E. (2005). Analysis of the statistical methods used to detect submaximal effort with the five-rung grip strength test. *Journal of Hand Therapy*, *18*, 10–18. doi:10.1197/j.jht.2004.10.004
- Shechtman, O., Hope, L. M., & Sindhu, B. S. (2007). Evaluation of the torque-velocity test of the BTE-Primus as a measure of sincerity of effort of grip strength. *Journal of Hand Therapy*, *20*, 326–334. doi:10.1197/j.jht.2007.07.009
- Shechtman, O., Sindhu, B. S., & Davenport, P. W. (2007). Using the force-time curve to detect maximal grip strength effort. *Journal of Hand Therapy*, *20*, 37–47, quiz 48. doi:10.1197/j.jht.2006.10.006
- Shechtman, O., & Taylor, C. (2000). The use of the rapid exchange grip test in detecting sincerity of effort, Part II: Validity of the test. *Journal of Hand Therapy*, *13*, 203–210.
- Shute, R. H., & Woodhouse, J. M. (1990). Visual fitness to drive after stroke or head injury. *Ophthalmic and Physiological Optics: The Journal of the British College of Ophthalmic Opticians*, *10*, 327–332.
- Staplin, L., Lococo, K., McKnight, A. J., McKnight, A. S., & Odenheimer, G. L. (1998). *Intersection negotiation problems of older drivers* (No. DTNH22–93–C–05237). Washington, DC: National Highway Traffic Safety Administration.
- 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.
- Stephens, B. W., McCarthy, D., Marsiske, M., Shechtman, O., Classen, S., Justiss, M., et al. (2005). International Older Driver Consensus Conference on assessment, remediation and counseling for transportation alternatives: Summary and recommendations. *Physical and Occupational Therapy in Geriatrics*, *23*, 103–121. doi:10.1300/J148v23n02_07
- Yale, S. H., Hansotia, P., Knapp, D., & Ehrfurth, J. (2003). Neurologic conditions: Assessing medical fitness to drive. *Clinical Medicine and Research*, *1*, 177–188. doi:10.3121/cmr.1.3.177